Testy dobré shody

Martina Litschmannová

VŠB TECHNICKÁ UNIVERZITA OSTRAVA | FAKULTA ELEKTROTECHNIKY A INFORMATIKY | KATEDRA APLIKOVANÉ MATEMATIKY

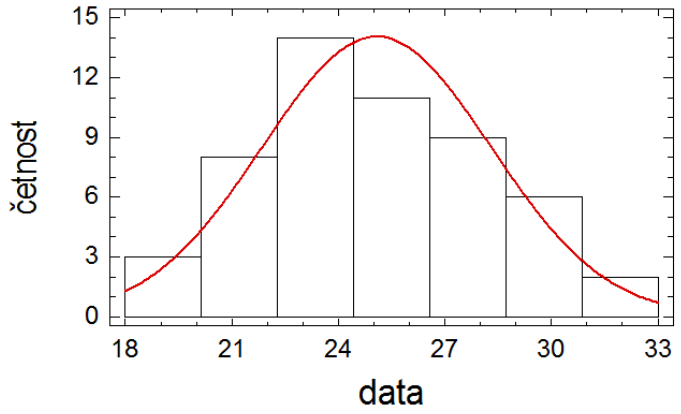
Obsah

- Grafická analýza pro ověření shody empirického a teoretického rozdělení
- χ^2 test dobré shody
  - ověření, zda jsou relativní četnosti jednotlivých variant rovny číslům π_{01}; π_{02}; …; π_{0k}
- χ^2 test dobré shody s očekávaným rozdělením
  - úplně specifikovaný test,
  - neúplně specifikovaných test,
- Kolmogorovův-Smirnovův jednovýběrový test
- Nejpoužívanější testy normality (přehled)

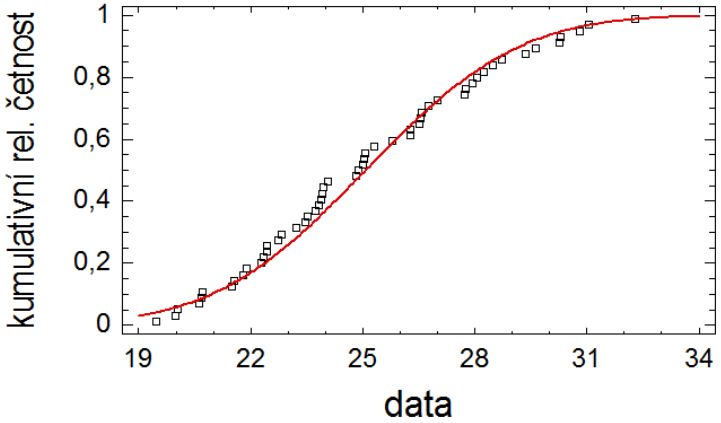
Poznámka: χ^2 čti: chí-kvadrát (česky) nebo „kaj skvére“ (anglicky)

Jak ověřit, zda se empirické rozdělení shoduje s teoretickým?

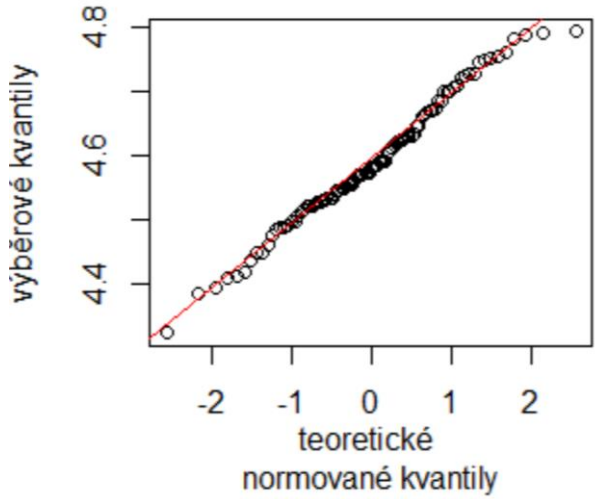
Srovnání histogramu
s teoretickou hustotou pravděpodobností



Srovnání kum. rel. četností (resp.
empirické distr. f-ce) s teoret. distr. f-cí



Q-Q graf



- Pomocí grafické analýzy můžeme metodou srovnání se stand. modely pouze odhadnout typ rozdělení!
- Objektivní míru shody dat s teoretickým modelem poskytují tzv. testy dobré shody.

Testy dobré shody

H_0: Data **jsou** výběrem z daného rozdělení. (Teoretické a výběrové rozdělení se **shodují**.)

H_A : Data **nejsou** výběrem z daného rozdělení. (Teoretické a výběrové rozdělení se **neshodují**.)

Vybrané testy dobré shody

- χ^2 test dobré shody - ověření, zda jsou rel. četnosti jednotlivých variant rovny číslům π_{01}; π_{02}; …; π_{0k},
- χ^2 - test dobré shody s očekávaným rozdělením
  - úplně specifikovaný test,
  - neúplně specifikovaných test,
- Kolmogorovův-Smirnovův jednovýběrový test

χ^2 test dobré shody - ověření, zda jsou rel. četnosti rovny číslům $\pi_{01}; \ldots; \pi_{0k}$

Motivační příklad

Bylo provedeno šetření mezi ženami staršími 15 let.

Mezi 246 náhodně oslovenými ženami bylo 80 (32,5 %) svobodných, 110 (44,7 %) vdaných, 30 (12,2 %) rozvedených a 26 (10,6 %) ovdovělých. Je známo (viz Český statistický úřad), že v ČR je mezi ženami staršími 15 let cca 24,8 % svobodných, 49,0 % vdaných, 12,6 % rozvedených a 13,6 % ovdovělých. Lze provedený výběr označit za reprezentativní?

Náznak řešení

Stav	svobodná	vdaná	rozvedená	ovdovělá	celkem
očekávaná rel. četnost π_{0_i}					

χ^2 test dobré shody - ověření, zda jsou rel. četnosti rovny číslům $\pi_{01}; \ldots; \pi_{0k}$

Motivační příklad

Bylo provedeno šetření mezi ženami staršími 15 let.

Mezi 246 náhodně oslovenými ženami bylo 80 (32,5 %) svobodných, 110 (44,7 %) vdaných, 30 (12,2 %) rozvedených a 26 (10,6 %) ovdovělých. Je známo (viz Český statistický úřad), že v ČR je mezi ženami staršími 15 let cca 24,8 % svobodných, 49,0 % vdaných, 12,6 % rozvedených a 13,6 % ovdovělých. Lze provedený výběr označit za reprezentativní?

Náznak řešení

Stav	svobodná	vdaná	rozvedená	ovdovělá	celkem
očekávaná rel. četnost π_{0_i}	0,248	0,490	0,126	0,136	**1,000**

χ^2 test dobré shody - ověření, zda jsou rel. četnosti rovny číslům $\pi_{01}; \ldots; \pi_{0k}$

Motivační příklad

Bylo provedeno šetření mezi ženami staršími 15 let.

Mezi 246 náhodně oslovenými ženami bylo 80 (32,5 %) svobodných, 110 (44,7 %) vdaných, 30 (12,2 %) rozvedených a 26 (10,6 %) ovdovělých. Je známo (viz Český statistický úřad), že v ČR je mezi ženami staršími 15 let cca 24,8 % svobodných, 49,0 % vdaných, 12,6 % rozvedených a 13,6 % ovdovělých. Lze provedený výběr označit za reprezentativní?

Náznak řešení

Stav	svobodná	vdaná	rozvedená	ovdovělá	celkem
očekávaná rel. četnost π_{0_i}	0,248	0,490	0,126	0,136	**1,000**
pozorovaná (angl. observed) četnost O_i					

χ^2 test dobré shody - ověření, zda jsou rel. četnosti rovny číslům $\pi_{01}; \ldots; \pi_{0k}$

Motivační příklad

Bylo provedeno šetření mezi ženami staršími 15 let.

Mezi 246 náhodně oslovenými ženami bylo 80 (32,5 %) svobodných, 110 (44,7 %) vdaných, 30 (12,2 %) rozvedených a 26 (10,6 %) ovdovělých. Je známo (viz Český statistický úřad), že v ČR je mezi ženami staršími 15 let cca 24,8 % svobodných, 49,0 % vdaných, 12,6 % rozvedených a 13,6 % ovdovělých. Lze provedený výběr označit za reprezentativní?

Náznak řešení

Stav	svobodná	vdaná	rozvedená	ovdovělá	celkem
očekávaná rel. četnost π_{0_i}	0,248	0,490	0,126	0,136	**1,000**
pozorovaná (angl. observed) četnost O_i	80	110	30	26	**256**

χ^2 test dobré shody - ověření, zda jsou rel. četnosti rovny číslům $\pi_{01}; \ldots; \pi_{0k}$

Motivační příklad

Bylo provedeno šetření mezi ženami staršími 15 let.

Mezi 246 náhodně oslovenými ženami bylo 80 (32,5 %) svobodných, 110 (44,7 %) vdaných, 30 (12,2 %) rozvedených a 26 (10,6 %) ovdovělých. Je známo (viz Český statistický úřad), že v ČR je mezi ženami staršími 15 let cca 24,8 % svobodných, 49,0 % vdaných, 12,6 % rozvedených a 13,6 % ovdovělých. Lze provedený výběr označit za reprezentativní?

Náznak řešení

Stav	svobodná	vdaná	rozvedená	ovdovělá	celkem
očekávaná rel. četnost π_{0_i}	0,248	0,490	0,126	0,136	**1,000**
pozorovaná (angl. observed) četnost O_i	80	110	30	26	**256**
očekávaná (angl. expected) četnost E_i					---

χ^2 test dobré shody - ověření, zda jsou rel. četnosti rovny číslům $\pi_{01}; \ldots; \pi_{0k}$

Motivační příklad

Bylo provedeno šetření mezi ženami staršími 15 let.

Mezi 246 náhodně oslovenými ženami bylo 80 (32,5 %) svobodných, 110 (44,7 %) vdaných, 30 (12,2 %) rozvedených a 26 (10,6 %) ovdovělých. Je známo (viz Český statistický úřad), že v ČR je mezi ženami staršími 15 let cca 24,8 % svobodných, 49,0 % vdaných, 12,6 % rozvedených a 13,6 % ovdovělých. Lze provedený výběr označit za reprezentativní?

Náznak řešení

Stav	svobodná	vdaná	rozvedená	ovdovělá	celkem
očekávaná rel. četnost π_{0_i}	0,248	0,490	0,126	0,136	**1,000**
pozorovaná (angl. observed) četnost O_i	80	110	30	26	**256**
očekávaná (angl. expected) četnost E_i	$0{,}248 \cdot 256$	$0{,}490 \cdot 256$	$0{,}126 \cdot 256$	$0{,}136 \cdot 256$	---

χ^2 test dobré shody - ověření, zda jsou rel. četnosti rovny číslům $\pi_{01}; \ldots; \pi_{0k}$

Motivační příklad

Bylo provedeno šetření mezi ženami staršími 15 let.

Mezi 246 náhodně oslovenými ženami bylo 80 (32,5 %) svobodných, 110 (44,7 %) vdaných, 30 (12,2 %) rozvedených a 26 (10,6 %) ovdovělých. Je známo (viz Český statistický úřad), že v ČR je mezi ženami staršími 15 let cca 24,8 % svobodných, 49,0 % vdaných, 12,6 % rozvedených a 13,6 % ovdovělých. Lze provedený výběr označit za reprezentativní?

Náznak řešení

Stav	svobodná	vdaná	rozvedená	ovdovělá	celkem
očekávaná rel. četnost π_{0_i}	0,248	0,490	0,126	0,136	**1,000**
pozorovaná (angl. observed) četnost O_i	80	110	30	26	**256**
očekávaná (angl. expected) četnost E_i	61,0	120,5	31,0	33,5	---

Liší se pozorované a očekávané četnosti statisticky významně?

χ^2 test dobré shody - ověření, zda jsou rel. četnosti rovny číslům $\pi_{01}; \ldots; \pi_{0k}$

V nejjednodušším případě použití χ^2 testu lze konečnou populaci roztřídit podle nějakého znaku do k disjunktních skupin (tzv. variant) a my chceme na základě náhodného výběru ověřit, zda jsou relativní četnosti jednotlivých variant rovny číslům $\pi_{01}; \pi_{02}; \ldots; \pi_{0k}$.

- **Předpoklady testu**: Dostatečně velký výběr. V praxi:
  - ✓ všechny očekávané četnosti větší nebo rovny 2 a
  - ✓ alespoň 80 % očekávaných četností větších než 5.

- **Testové kritérium**: $G = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \sim \chi^2\,(k-1)$

- $\boldsymbol{p - hodnota} = 1 - F_0(x_{OBS})$

χ^2 test dobré shody - ověření, zda jsou rel. četnosti rovny číslům $\pi_{01}; \ldots; \pi_{0k}$

Motivační příklad

Bylo provedeno šetření mezi ženami staršími 15 let.

Mezi 246 náhodně oslovenými ženami bylo 80 (32,5 %) svobodných, 110 (44,7 %) vdaných, 30 (12,2 %) rozvedených a 26 (10,6 %) ovdovělých. Je známo (viz Český statistický úřad), že v ČR je mezi ženami staršími 15 let cca 24,8 % svobodných, 49,0 % vdaných, 12,6 % rozvedených a 13,6 % ovdovělých. Lze provedený výběr označit za reprezentativní?

Řešení

H_0: Provedený výběr **je** výběrem z populace, v níž jsou relativní četnosti jednotlivých variant dány tabulkou:

Stav	svobodná	vdaná	rozvedená	ovdovělá	celkem
očekávaná rel. četnost π_{0_i}	0,248	0,490	0,126	0,136	**1,000**

H_A: $\neg H_0$, tj. provedený výběr není výběrem z populace, v níž jsou relativní četnosti jednotlivých variant dány výše uvedenou tabulkou.

χ^2 test dobré shody - ověření, zda jsou rel. četnosti rovny číslům $\pi_{01}; \ldots; \pi_{0k}$

Motivační příklad

Bylo provedeno šetření mezi ženami staršími 15 let.

Mezi 246 náhodně oslovenými ženami bylo 80 (32,5 %) svobodných, 110 (44,7 %) vdaných, 30 (12,2 %) rozvedených a 26 (10,6 %) ovdovělých. Je známo (viz Český statistický úřad), že v ČR je mezi ženami staršími 15 let cca 24,8 % svobodných, 49,0 % vdaných, 12,6 % rozvedených a 13,6 % ovdovělých. Lze provedený výběr označit za reprezentativní?

Řešení

Stav	svobodná	vdaná	rozvedená	ovdovělá	celkem
očekávaná rel. četnost π_{0_i}	0,248	0,490	0,126	0,136	**1,000**
pozorovaná (angl. observed) četnost O_i	80	110	30	26	**256**

Ověření předpokladů:

χ^2 test dobré shody - ověření, zda jsou rel. četnosti rovny číslům $\pi_{01}; \ldots; \pi_{0k}$

Motivační příklad

Bylo provedeno šetření mezi ženami staršími 15 let.

Mezi 246 náhodně oslovenými ženami bylo 80 (32,5 %) svobodných, 110 (44,7 %) vdaných, 30 (12,2 %) rozvedených a 26 (10,6 %) ovdovělých. Je známo (viz Český statistický úřad), že v ČR je mezi ženami staršími 15 let cca 24,8 % svobodných, 49,0 % vdaných, 12,6 % rozvedených a 13,6 % ovdovělých. Lze provedený výběr označit za reprezentativní?

Řešení

Stav	svobodná	vdaná	rozvedená	ovdovělá	celkem
očekávaná rel. četnost π_{0_i}	0,248	0,490	0,126	0,136	**1,000**
pozorovaná (angl. observed) četnost O_i	80	110	30	26	**256**
očekávaná (angl. expected) četnost E_i	61,0	120,5	31,0	33,5	---

Ověření předpokladů: Všechny očekávané četnosti jsou větší než 5.

Pozorovaná hodnota: $x_{OBS} = \sum_{i=1}^{4} \frac{(O_i - E_i)^2}{E_i} = \frac{(80-61{,}0)^2}{61{,}0} + \frac{(110-120{,}5)^2}{120{,}5} + \frac{(30-31{,}0)^2}{31{,}0} + \frac{(26-33{,}5)^2}{33{,}5} = 8{,}53$

χ^2 test dobré shody - ověření, zda jsou rel. četnosti rovny číslům $\pi_{01}; \ldots; \pi_{0k}$

Motivační příklad

Bylo provedeno šetření mezi ženami staršími 15 let.

Mezi 246 náhodně oslovenými ženami bylo 80 (32,5 %) svobodných, 110 (44,7 %) vdaných, 30 (12,2 %) rozvedených a 26 (10,6 %) ovdovělých. Je známo (viz Český statistický úřad), že v ČR je mezi ženami staršími 15 let cca 24,8 % svobodných, 49,0 % vdaných, 12,6 % rozvedených a 13,6 % ovdovělých. Lze provedený výběr označit za reprezentativní?

Řešení

Stav	svobodná	vdaná	rozvedená	ovdovělá	celkem
očekávaná rel. četnost π_{0_i}	0,248	0,490	0,126	0,136	**1,000**
pozorovaná (angl. observed) četnost O_i	80	110	30	26	**256**
očekávaná (angl. expected) četnost E_i	61,0	120,5	31,0	33,5	---

Pozorovaná hodnota: $x_{OBS} = 8{,}53$

p-hodnota: $p\text{–}hodnota = 1 - F_0(x_{OBS})$, kde $F_0(x)$ je dist. funkce χ^2 rozdělení s 3 $(= 4 - 1)$ st. volnosti,
$p\text{–}hodnota = 1 - F_0(8{,}53) = 0{,}036$. (*1-pchisq(8.53,3)*)

χ^2 test dobré shody - ověření, zda jsou rel. četnosti rovny číslům $\pi_{01}; \ldots; \pi_{0k}$

Motivační příklad

Bylo provedeno šetření mezi ženami staršími 15 let.

Mezi 246 náhodně oslovenými ženami bylo 80 (32,5 %) svobodných, 110 (44,7 %) vdaných, 30 (12,2 %) rozvedených a 26 (10,6 %) ovdovělých. Je známo (viz Český statistický úřad), že v ČR je mezi ženami staršími 15 let cca 24,8 % svobodných, 49,0 % vdaných, 12,6 % rozvedených a 13,6 % ovdovělých. Lze provedený výběr označit za reprezentativní?

Řešení

Stav	svobodná	vdaná	rozvedená	ovdovělá	celkem
očekávaná rel. četnost π_{0_i}	0,248	0,490	0,126	0,136	**1,000**
pozorovaná (angl. observed) četnost O_i	80	110	30	26	**256**
očekávaná (angl. expected) četnost E_i	61,0	120,5	31,0	33,5	---

p-hodnota: $p\text{–}hodnota = 0{,}036$

Rozhodnutí: Na hladině významnosti 0,05 zamítáme H_0, tj. výběr **nelze** považovat za reprezentativní.

χ^2 test dobré shody s očekávaným rozdělením

H_0: Data **jsou** výběrem z daného rozdělení. (Teoretické a výběrové rozdělení se **shodují**.)

H_A : Data **nejsou** výběrem z daného rozdělení. (Teoretické a výběrové rozdělení se **neshodují**.)

Rozlišujeme:

- **Úplně specifikovaný test** – v nulové hypotéze jsou specifikovány všechny parametry teoretického rozdělení (např. H_0: Výběr pochází z Poissonova rozdělení se střední hodnotou 12.)

- **Neúplně specifikovaný test** – v nulové hypotéze nejsou specifikovány všechny parametry teoretického rozdělení (např. H_0: Výběr pochází z Poissonova rozdělení.)

  V tomto případě musíme nespecifikované parametry teoretického rozdělení odhadnout (bodový odhad) z výběru.

χ^2 test dobré shody s očekávaným rozdělením

H_0: Data **jsou** výběrem z daného rozdělení. (Teoretické a výběrové rozdělení se **shodují**.)

H_A : Data **nejsou** výběrem z daného rozdělení. (Teoretické a výběrové rozdělení se **neshodují**.)

- **Předpoklady testu**: Dostatečně velký výběr – v praxi:
  - ✓ všechny očekávané četnosti větší nebo rovny 2 a
  - ✓ alespoň 80 % očekávaných četností větších než 5.

- **Testové kritérium**: $G = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \sim \chi^2\ (k-1-h)$, kde h je počet odhadovaných parametrů teoretického rozdělení

- $\boldsymbol{p - hodnota} = 1 - F_0(x_{OBS})$

χ^2 test dobré shody s očekávaným rozdělením

Je-li teoretické rozdělení **diskrétní**, pak:

1) stanovíme nulovou a alternativní hypotézu,
2) označíme pozorované varianty NV x_i, kde $i = 1, 2, \dots, k$,
3) určíme pozorované četnosti O_i jednotlivých variant NV $\left(\sum_{i=1}^{k} O_i = n\right)$,
4) určíme očekávané rel. četnosti π_i jednotlivých variant NV $\left(\pi_i = P_0(x_i)\right)$,
5) určíme očekávané četnosti E_i jednotlivých variant NV $(E_i = n\pi_i)$,
6) ověříme předpoklady testu (očekávané četnosti větší nebo rovno 2, alespoň 80 % očekávaných četností větší než 5),
7) určíme pozorovanou hodnotu testové statistiky $\left(x_{OBS} = \sum_{i=1}^{4} \frac{(O_i - E_i)^2}{E_i}\right)$,
8) určíme *p*-hodnotu ($p\text{–}hodnota = 1 - F_0(x_{OBS})$, kde $F_0(x)$ je distribuční funkce χ^2 rozdělení s $k - h - 1$ stupni volnosti, kde *h* je počet odhadovaných parametrů teoretického rozdělení),
9) rozhodneme o výsledku testu.

χ^2 test dobré shody s očekávaným rozdělením

Výrobní firma odhaduje počet poruch určitého zařízení během dne pomocí Poissonova rozdělení se střední hodnotou 1,2. Zaměstnanci zaznamenali pro kontrolu skutečné počty poruch celkem ve 150 dnech (výsledky jsou uvedeny v níže uvedené tabulce). Ověřte čistým testem významnosti, zda lze počet poruch daného zařízení během dne skutečně modelovat pomocí Poissonova rozdělení s parametrem λt=1,2.

x_i – počet poruch během dne	0	1	2	3	4 a více	celkem
O_i – počet dní, v nichž byl pozorován počet poruch x_i	52	48	36	10	4	150

Řešení:

H_0: Počet poruch daného zařízení během jednoho dne (náhodná veličina X) má Poissonovo rozdělení s parametrem $\lambda t = 1{,}2$, neboli výběr pochází z Poissonova rozdělení s parametrem $\lambda t = 1{,}2$.

H_A: $\neg H_0$

χ^2 test dobré shody s očekávaným rozdělením

Výrobní firma odhaduje počet poruch určitého zařízení během dne pomocí Poissonova rozdělení se střední hodnotou 1,2. Zaměstnanci zaznamenali pro kontrolu skutečné počty poruch celkem ve 150 dnech (výsledky jsou uvedeny v níže uvedené tabulce). Ověřte čistým testem významnosti, zda lze počet poruch daného zařízení během dne skutečně modelovat pomocí Poissonova rozdělení s parametrem λt=1,2.

x_i – počet poruch během dne	0	1	2	3	4 a více	celkem
O_i – počet dní, v nichž byl pozorován počet poruch x_i	52	48	36	10	4	150
π_i - očekávané pravděpodobnosti výskytu						

Řešení:

H_0: Počet poruch daného zařízení během jednoho dne (náhodná veličina X) má Poissonovo rozdělení s parametrem $\lambda t = 1{,}2$, neboli výběr pochází z Poissonova rozdělení s parametrem $\lambda t = 1{,}2$.

H_A: $\neg H_0$

χ^2 test dobré shody s očekávaným rozdělením

Výrobní firma odhaduje počet poruch určitého zařízení během dne pomocí Poissonova rozdělení se střední hodnotou 1,2. Zaměstnanci zaznamenali pro kontrolu skutečné počty poruch celkem ve 150 dnech (výsledky jsou uvedeny v níže uvedené tabulce). Ověřte čistým testem významnosti, zda lze počet poruch daného zařízení během dne skutečně modelovat pomocí Poissonova rozdělení s parametrem λt=1,2.

x_i – počet poruch během dne	0	1	2	3	4 a více	celkem
O_i – počet dní, v nichž byl pozorován počet poruch x_i	52	48	36	10	4	150
π_i - očekávané pravděpodobnosti výskytu	$P(X=0)$	$P(X=1)$	$P(X=2)$	$P(X=3)$	$P_0(X \geq 4)$	1,000

X … počet poruch během dne; $X \sim Po(\lambda t = 1{,}2)$

Řešení:

H_0: Počet poruch daného zařízení během jednoho dne (náhodná veličina X) má Poissonovo rozdělení s parametrem $\lambda t = 1{,}2$, neboli výběr pochází z Poissonova rozdělení s parametrem $\lambda t = 1{,}2$.

H_A: $\neg H_0$

χ^2 test dobré shody s očekávaným rozdělením

Výrobní firma odhaduje počet poruch určitého zařízení během dne pomocí Poissonova rozdělení se střední hodnotou 1,2. Zaměstnanci zaznamenali pro kontrolu skutečné počty poruch celkem ve 150 dnech (výsledky jsou uvedeny v níže uvedené tabulce). Ověřte čistým testem významnosti, zda lze počet poruch daného zařízení během dne skutečně modelovat pomocí Poissonova rozdělení s parametrem λt=1,2.

x_i – počet poruch během dne	0	1	2	3	4 a více	celkem
O_i – počet dní, v nichž byl pozorován počet poruch x_i	52	48	36	10	4	150
π_i - očekávané pravděpodobnosti výskytu	0,301	0,361	0,217	0,087	0,034	1,000

Řešení:

H_0: Počet poruch daného zařízení během jednoho dne (náhodná veličina X) má Poissonovo rozdělení s parametrem $\lambda t = 1{,}2$, neboli výběr pochází z Poissonova rozdělení s parametrem $\lambda t = 1{,}2$.

H_A: $\neg H_0$

χ^2 test dobré shody s očekávaným rozdělením

Výrobní firma odhaduje počet poruch určitého zařízení během dne pomocí Poissonova rozdělení se střední hodnotou 1,2. Zaměstnanci zaznamenali pro kontrolu skutečné počty poruch celkem ve 150 dnech (výsledky jsou uvedeny v níže uvedené tabulce). Ověřte čistým testem významnosti, zda lze počet poruch daného zařízení během dne skutečně modelovat pomocí Poissonova rozdělení s parametrem λt=1,2.

x_i – počet poruch během dne	0	1	2	3	4 a více	celkem
O_i – počet dní, v nichž byl pozorován počet poruch x_i	52	48	36	10	4	150
π_i - očekávané pravděpodobnosti výskytu	0,301	0,361	0,217	0,087	0,034	1,000
E_i - očekávané četnosti výskytu						---

Řešení:

H_0: Počet poruch daného zařízení během jednoho dne (náhodná veličina X) má Poissonovo rozdělení s parametrem $\lambda t = 1{,}2$, neboli výběr pochází z Poissonova rozdělení s parametrem $\lambda t = 1{,}2$.

H_A: $\neg H_0$

χ^2 test dobré shody s očekávaným rozdělením

Výrobní firma odhaduje počet poruch určitého zařízení během dne pomocí Poissonova rozdělení se střední hodnotou 1,2. Zaměstnanci zaznamenali pro kontrolu skutečné počty poruch celkem ve 150 dnech (výsledky jsou uvedeny v níže uvedené tabulce). Ověřte čistým testem významnosti, zda lze počet poruch daného zařízení během dne skutečně modelovat pomocí Poissonova rozdělení s parametrem λt=1,2.

x_i – počet poruch během dne	0	1	2	3	4 a více	celkem
O_i – počet dní, v nichž byl pozorován počet poruch x_i	52	48	36	10	4	150
π_i - očekávané pravděpodobnosti výskytu	0,301	0,361	0,217	0,087	0,034	1,000
E_i - očekávané četnosti výskytu	45,2	54,2	32,6	13,1	5,1	---

Řešení:

H_0: Počet poruch daného zařízení během jednoho dne (náhodná veličina X) má Poissonovo rozdělení s parametrem $\lambda t = 1{,}2$, neboli výběr pochází z Poissonova rozdělení s parametrem $\lambda t = 1{,}2$.

H_A: $\neg H_0$

χ^2 test dobré shody s očekávaným rozdělením

Výrobní firma odhaduje počet poruch určitého zařízení během dne pomocí Poissonova rozdělení se střední hodnotou 1,2. Zaměstnanci zaznamenali pro kontrolu skutečné počty poruch celkem ve 150 dnech (výsledky jsou uvedeny v níže uvedené tabulce). Ověřte čistým testem významnosti, zda lze počet poruch daného zařízení během dne skutečně modelovat pomocí Poissonova rozdělení s parametrem λt=1,2.

x_i – počet poruch během dne	0	1	2	3	4 a více	celkem
O_i – počet dní, v nichž byl pozorován počet poruch x_i	52	48	36	10	4	150
π_i - očekávané pravděpodobnosti výskytu	0,301	0,361	0,217	0,087	0,034	1,000
E_i - očekávané četnosti výskytu	45,2	54,2	32,6	13,1	5,1	---

$= 150 \cdot 0{,}301$

Řešení:

H_0: Počet poruch daného zařízení během jednoho dne (náhodná veličina X) má Poissonovo rozdělení s parametrem $\lambda t = 1{,}2$, neboli výběr pochází z Poissonova rozdělení s parametrem $\lambda t = 1{,}2$.

H_A: $\neg H_0$

χ^2 test dobré shody s očekávaným rozdělením

Výrobní firma odhaduje počet poruch určitého zařízení během dne pomocí Poissonova rozdělení se střední hodnotou 1,2. Zaměstnanci zaznamenali pro kontrolu skutečné počty poruch celkem ve 150 dnech (výsledky jsou uvedeny v níže uvedené tabulce). Ověřte čistým testem významnosti, zda lze počet poruch daného zařízení během dne skutečně modelovat pomocí Poissonova rozdělení s parametrem λt=1,2.

x_i – počet poruch během dne	0	1	2	3	4 a více	celkem
O_i – počet dní, v nichž byl pozorován počet poruch x_i	52	48	36	10	4	150
π_i - očekávané pravděpodobnosti výskytu	0,301	0,361	0,217	0,087	0,034	1,000
E_i - očekávané četnosti výskytu	45,2	54,2	32,6	13,1	5,1	---

$= 150 \cdot 0{,}361$

Řešení:

H_0: Počet poruch daného zařízení během jednoho dne (náhodná veličina X) má Poissonovo rozdělení s parametrem $\lambda t = 1{,}2$, neboli výběr pochází z Poissonova rozdělení s parametrem $\lambda t = 1{,}2$.

H_A: $\neg H_0$

χ^2 test dobré shody s očekávaným rozdělením

Výrobní firma odhaduje počet poruch určitého zařízení během dne pomocí Poissonova rozdělení se střední hodnotou 1,2. Zaměstnanci zaznamenali pro kontrolu skutečné počty poruch celkem ve 150 dnech (výsledky jsou uvedeny v níže uvedené tabulce). Ověřte čistým testem významnosti, zda lze počet poruch daného zařízení během dne skutečně modelovat pomocí Poissonova rozdělení s parametrem λt=1,2.

x_i – počet poruch během dne	0	1	2	3	4 a více	celkem
O_i – počet dní, v nichž byl pozorován počet poruch x_i	52	48	36	10	4	150
π_i - očekávané pravděpodobnosti výskytu	0,301	0,361	0,217	0,087	0,034	1,000
E_i - očekávané četnosti výskytu	45,2	54,2	32,6	13,1	5,1	---

Řešení:

H_0: Počet poruch daného zařízení během jednoho dne (náhodná veličina X) má Poissonovo rozdělení s parametrem $\lambda t = 1{,}2$, neboli výběr pochází z Poissonova rozdělení s parametrem $\lambda t = 1{,}2$.

H_A: $\neg H_0$

χ^2 test dobré shody s očekávaným rozdělením

Výrobní firma odhaduje počet poruch určitého zařízení během dne pomocí Poissonova rozdělení se střední hodnotou 1,2. Zaměstnanci zaznamenali pro kontrolu skutečné počty poruch celkem ve 150 dnech (výsledky jsou uvedeny v níže uvedené tabulce). Ověřte čistým testem významnosti, zda lze počet poruch daného zařízení během dne skutečně modelovat pomocí Poissonova rozdělení s parametrem λt=1,2.

x_i – počet poruch během dne	0	1	2	3	4 a více	celkem
O_i – počet dní, v nichž byl pozorován počet poruch x_i	52	48	36	10	4	150
π_i - očekávané pravděpodobnosti výskytu	0,301	0,361	0,217	0,087	0,034	1,000
E_i - očekávané četnosti výskytu	45,2	54,2	32,6	13,1	5,1	---

Řešení:

Pozorovaná hodnota: $x_{OBS} = \sum_{i=1}^{5} \frac{(O_i - E_i)^2}{E_i} = \frac{(52-45{,}2)^2}{45{,}2} + \cdots + \frac{(4-5{,}1)^2}{5{,}1} = 3{,}13.$

Testové kritérium má za předpokladu platnosti H_0 χ^2 rozdělení s 4 (k−1−h) stupni volnosti. (Počet variant k=5, počet odhadovaných parametrů h=0.)

χ^2 test dobré shody s očekávaným rozdělením

Výrobní firma odhaduje počet poruch určitého zařízení během dne pomocí Poissonova rozdělení se střední hodnotou 1,2. Zaměstnanci zaznamenali pro kontrolu skutečné počty poruch celkem ve 150 dnech (výsledky jsou uvedeny v níže uvedené tabulce). Ověřte čistým testem významnosti, zda lze počet poruch daného zařízení během dne skutečně modelovat pomocí Poissonova rozdělení s parametrem λt=1,2.

x_i – počet poruch během dne	0	1	2	3	4 a více	celkem
O_i – počet dní, v nichž byl pozorován počet poruch x_i	52	48	36	10	4	150
π_i - očekávané pravděpodobnosti výskytu	0,301	0,361	0,217	0,087	0,034	1,000
E_i - očekávané četnosti výskytu	45,2	54,2	32,6	13,1	5,1	---

Řešení:

Pozorovaná hodnota: $x_{OBS} = \sum_{i=1}^{5} \frac{(O_i - E_i)^2}{E_i} = \frac{(52-45{,}2)^2}{45{,}2} + \cdots + \frac{(4-5{,}1)^2}{5{,}1} = 3{,}13.$

p-hodnota: $p\text{–}hodnota = 1 - F_0(x_{OBS})$, kde $F_0(x)$ je distribuční funkce χ^2 rozdělení s 4 stupni volnosti.

$$p\text{–}hodnota = 1 - F_0(3{,}13) = 0{,}540$$ (*1-pchisq(3.13,4)*)

χ^2 test dobré shody s očekávaným rozdělením

Výrobní firma odhaduje počet poruch určitého zařízení během dne pomocí Poissonova rozdělení se střední hodnotou 1,2. Zaměstnanci zaznamenali pro kontrolu skutečné počty poruch celkem ve 150 dnech (výsledky jsou uvedeny v níže uvedené tabulce). Ověřte čistým testem významnosti, zda lze počet poruch daného zařízení během dne skutečně modelovat pomocí Poissonova rozdělení s parametrem λt=1,2.

x_i – počet poruch během dne	0	1	2	3	4 a více	celkem
O_i – počet dní, v nichž byl pozorován počet poruch x_i	52	48	36	10	4	150
π_i - očekávané pravděpodobnosti výskytu	0,301	0,361	0,217	0,087	0,034	1,000
E_i - očekávané četnosti výskytu	45,2	54,2	32,6	13,1	5,1	---

Řešení:

$$p\text{–}hodnota = 1 - F_0(3{,}13) = 0{,}540$$

Rozhodnutí:

Na hladině významnosti 0,05 nezamítáme nulovou hypotézu, tzn. nemáme námitek proti použití Poissonova rozdělení s parametrem 1,2 pro odhad počtu poruch daného zařízení během jednoho dne.

χ^2 test dobré shody s očekávaným rozdělením

Je-li teoretické rozdělení spojité, pak:

1) stanovíme nulovou a alternativní hypotézu,
2) pozorované hodnoty NV roztřídíme do *k* tříd (obvykle volíme $5 \leq k \leq 15$),
3) určíme pozorované četnosti O_i v jednotlivých třídách $\left(\sum_{i=1}^{k} O_i = n\right)$,
4) určíme očekávané rel. četnosti π_i v jednotlivých třídách,
5) určíme očekávané četnosti E_i v jednotlivých třídách $(E_i = n\pi_i)$,
6) ověříme předpoklady testu (očekávané četnosti větší nebo rovno 2, alespoň 80 % očekávaných četností větší než 5),
7) není-li předpoklad testu splněn, pokusíme se třídy korigovat (nemusí být stejně velké),
8) určíme pozorovanou hodnotu testové statistiky $\left(x_{OBS} = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}\right)$,
9) určíme *p*-hodnotu ($p\text{–}hodnota = 1 - F_0(x_{OBS})$, kde $F_0(x)$ je distribuční funkce χ^2 rozdělení s $k - h - 1$ stupni volnosti, kde *h* je počet odhadovaných parametrů teoretického rozdělení),
10) rozhodneme o výsledku testu.

χ^2 test dobré shody s očekávaným rozdělením

Na dálnici byly v průběhu několika minut měřeny časové odstupy (s) mezi průjezdy jednotlivých vozidel. Zjištěné hodnoty těchto odstupů jsou uvedeny v tabulce:

2,5	6,8	5,0	9,8	4,0	2,3	4,2	19	8,7	7,7	5,9	5,3	8,4	3,6	9,2
4,3	2,6	13,0	5,4	8,6	4,2	2,9	1,5	1,8	1,6	5,9	8,3	5,2	6,9	5,1
1,3	6,4	6,5	5,7	3,6	4,8	4,0	7,3	24,9	10,6	15,0	5,3	4,0	3,3	6,0
4,6	1,6	1,9	1,5	11,1	4,3	5,5	2,1	2,9	3,0	3,8	1,0	1,5	8,6	4,4
6,8	5,2	3,0	8,0	4,0	4,7	7,3	2,3	1,9	1,9	4,6	6,4	5,3	3,9	2,4
1,2	6,2	4,3	2,6	2,7	2,0	0,8	3,7	6,9	2,8	4,3	4,9	4,1	4,5	4,4
11,9	9,0	5,6	4,8	2,8	2,1	4,3	1,0	1,6	2,5	2,2	1,3	1,8	1,6	3,8
3,1	1,6	4,9	1,8	3,9	3,4	1,6	4,5	5,8	6,9	1,8	2,6	6,8	2,5	1,9
3,1	10,8	1,6	2,0	4,9	11,2	1,6	2,2	3,8	1,1	1,8	1,4			

Ověřte čistým testem významnosti, zda lze časové odstupy mezi vozidly modelovat pomocí náhodné veličinu s norm. rozdělením.

Řešení: X … časový odstup mezi průjezdy jednotlivých vozidel.

H_0: Časové odstupy mezi průjezdy jednotlivých vozidel mají normální rozdělení.

H_A: Časové odstupy mezi průjezdy jednotlivých vozidel nemají normální rozdělení.

χ^2 test dobré shody s očekávaným rozdělením

Řešení:
X … časový odstup mezi průjezdy jednotlivých vozidel.

H_0: Časové odstupy mezi průjezdy jednotlivých vozidel mají normální rozdělení.

H_A: Časové odstupy mezi průjezdy jednotlivých vozidel nemají normální rozdělení.

- Neúplně specifikovaný test – normální rozdělení má 2 parametry (μ a σ^2), které je třeba odhadnout.

$$\hat{\mu} = \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{\sum_{i=1}^{132} x_i}{132} = 4{,}6, \quad \hat{\sigma}^2 = s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} = \frac{\sum_{i=1}^{132}(x_i - 4{,}6)^2}{131} = 10{,}9$$

Proč musíme nyní pozorované hodnoty kategorizovat ?

Pravděpodobnostní f-ce spojité NV je nulová. Pravděpodobnost výskytu spojité NV na určitém intervalu již nulová není!

χ^2 test dobré shody s očekávaným rozdělením

Řešení:

- Definiční obor náhodné veličiny rozdělíme například do 13 třídících intervalů, tj. náhodnou veličinu **kategorizujeme**.

i	Třídící interval (s)	Empirické četnosti O_i
1	$(-\infty; 1{,}5\rangle$	11
2	$(1{,}5; 1{,}8\rangle$	13
3	$(1{,}8; 2{,}0\rangle$	7
4	$(2{,}0; 2{,}5\rangle$	10
5	$(2{,}5; 2{,}9\rangle$	8
6	$(2{,}9; 3{,}6\rangle$	8
7	$(3{,}6; 4{,}0\rangle$	10
8	$(4{,}0; 4{,}4\rangle$	10
9	$(4{,}4; 4{,}9\rangle$	10
10	$(4{,}9; 5{,}8\rangle$	12
11	$(5{,}8; 6{,}8\rangle$	10
12	$(6{,}8; 8{,}7\rangle$	12
13	$(8{,}7; \infty)$	11
Celkem	-	132

χ^2 test dobré shody s očekávaným rozdělením

Řešení:

- Určíme očekávané pravděpodobnosti.

i	Třídící interval (s)	Empirické četnosti O_i	Očekávané pravd. π_{0_i}
1	$(-\infty; 1{,}5\rangle$	11	
2	$(1{,}5; 1{,}8\rangle$	13	
3	$(1{,}8; 2{,}0\rangle$	7	
4	$(2{,}0; 2{,}5\rangle$	10	
5	$(2{,}5; 2{,}9\rangle$	8	
6	$(2{,}9; 3{,}6\rangle$	8	
7	$(3{,}6; 4{,}0\rangle$	10	
8	$(4{,}0; 4{,}4\rangle$	10	
9	$(4{,}4; 4{,}9\rangle$	10	
10	$(4{,}9; 5{,}8\rangle$	12	
11	$(5{,}8; 6{,}8\rangle$	10	
12	$(6{,}8; 8{,}7\rangle$	12	
13	$(8{,}7; \infty)$	11	
Celkem	-	132	1,000

χ^2 test dobré shody s očekávaným rozdělením

Řešení:

- Určíme očekávané pravděpodobnosti.

i	Třídící interval (s)	Empirické četnosti O_i	Očekávané pravd. π_{0_i}
1	$(-\infty; 1{,}5\rangle$	11	0,174
2	$(1{,}5; 1{,}8\rangle$	13	0,024
3	$(1{,}8; 2{,}0\rangle$	7	0,017
4	$(2{,}0; 2{,}5\rangle$	10	0,047
5	$(2{,}5; 2{,}9\rangle$	8	0,041
6	$(2{,}9; 3{,}6\rangle$	8	0,078
7	$(3{,}6; 4{,}0\rangle$	10	0,047
8	$(4{,}0; 4{,}4\rangle$	10	0,048
9	$(4{,}4; 4{,}9\rangle$	10	0,060
10	$(4{,}9; 5{,}8\rangle$	12	0,106
11	$(5{,}8; 6{,}8\rangle$	10	0,106
12	$(6{,}8; 8{,}7\rangle$	12	0,145
13	$(8{,}7; \infty)$	11	0,107
Celkem	-	132	1,000

$P(X \in (-\infty; 1{,}5\rangle\ ) = P(X < 1{,}5) = F(1{,}5)$, kde $F(x)$ je distr. f-ce rozdělení $N(4{,}6; 10{,}9)$

χ^2 test dobré shody s očekávaným rozdělením

Řešení:

- Určíme očekávané pravděpodobnosti.

i	Třídící interval (s)	Empirické četnosti O_i	Očekávané pravd. π_{0_i}
1	$(-\infty; 1{,}5\rangle$	11	0,174
2	$(1{,}5; 1{,}8\rangle$	13	0,024
3	$(1{,}8; 2{,}0\rangle$	7	0,017
4	$(2{,}0; 2{,}5\rangle$	10	0,047
5	$(2{,}5; 2{,}9\rangle$	8	0,041
6	$(2{,}9; 3{,}6\rangle$	8	0,078
7	$(3{,}6; 4{,}0\rangle$	10	0,047
8	$(4{,}0; 4{,}4\rangle$	10	0,048
9	$(4{,}4; 4{,}9\rangle$	10	0,060
10	$(4{,}9; 5{,}8\rangle$	12	0,106
11	$(5{,}8; 6{,}8\rangle$	10	0,106
12	$(6{,}8; 8{,}7\rangle$	12	0,145
13	$(8{,}7; \infty)$	11	0,107
Celkem	-	132	1,000

$P(X \in (1{,}5; 1{,}8\rangle\ ) = F(1{,}8) - F(1{,}5)$, kde $F(x)$ je distr. f-ce rozdělení $N(4{,}6; 10{,}9)$

χ^2 test dobré shody s očekávaným rozdělením

Řešení:

- Určíme očekávané pravděpodobnosti.

i	Třídící interval (s)	Empirické četnosti O_i	Očekávané pravd. π_{0_i}
1	$(-\infty; 1{,}5\rangle$	11	0,174
2	$(1{,}5; 1{,}8\rangle$	13	0,024
3	$(1{,}8; 2{,}0\rangle$	7	0,017
4	$(2{,}0; 2{,}5\rangle$	10	0,047
5	$(2{,}5; 2{,}9\rangle$	8	0,041
6	$(2{,}9; 3{,}6\rangle$	8	0,078
7	$(3{,}6; 4{,}0\rangle$	10	0,047
8	$(4{,}0; 4{,}4\rangle$	10	0,048
9	$(4{,}4; 4{,}9\rangle$	10	0,060
10	$(4{,}9; 5{,}8\rangle$	12	0,106
11	$(5{,}8; 6{,}8\rangle$	10	0,106
12	$(6{,}8; 8{,}7\rangle$	12	0,145
13	$(8{,}7; \infty)$	11	0,107
Celkem	-	132	1,000

$P(X \in (1{,}8; 2{,}0\rangle\ ) = F(2{,}0) - F(1{,}8)$, kde $F(x)$ je distr. f-ce rozdělení $N(4{,}6; 10{,}9)$

χ^2 test dobré shody s očekávaným rozdělením

Řešení:

- Určíme očekávané pravděpodobnosti.

i	Třídící interval (s)	Empirické četnosti O_i	Očekávané pravd. π_{0_i}
1	$(-\infty; 1{,}5\rangle$	11	0,174
2	$(1{,}5; 1{,}8\rangle$	13	0,024
3	$(1{,}8; 2{,}0\rangle$	7	0,017
4	$(2{,}0; 2{,}5\rangle$	10	0,047
5	$(2{,}5; 2{,}9\rangle$	8	0,041
6	$(2{,}9; 3{,}6\rangle$	8	0,078
7	$(3{,}6; 4{,}0\rangle$	10	0,047
8	$(4{,}0; 4{,}4\rangle$	10	0,048
9	$(4{,}4; 4{,}9\rangle$	10	0,060
10	$(4{,}9; 5{,}8\rangle$	12	0,106
11	$(5{,}8; 6{,}8\rangle$	10	0,106
12	$(6{,}8; 8{,}7\rangle$	12	0,145
13	$(8{,}7; \infty)$	11	0,107
Celkem	-	132	1,000

$P(X \in (8{,}7; \infty)\ ) = P(X > 8{,}7) = 1 - F(8{,}7)$, kde $F(x)$ je distr. f-ce rozdělení $N(4{,}6; 10{,}9)$

χ^2 test dobré shody s očekávaným rozdělením

Řešení:

- Určíme očekávané četnosti.

i	Třídící interval (s)	Empirické četnosti O_i	Očekávané pravd. π_{0_i}	Očekávané četnosti E_i
1	$(-\infty; 1{,}5\rangle$	11	0,174	
2	$(1{,}5; 1{,}8\rangle$	13	0,024	
3	$(1{,}8; 2{,}0\rangle$	7	0,017	
4	$(2{,}0; 2{,}5\rangle$	10	0,047	
5	$(2{,}5; 2{,}9\rangle$	8	0,041	
6	$(2{,}9; 3{,}6\rangle$	8	0,078	
7	$(3{,}6; 4{,}0\rangle$	10	0,047	
8	$(4{,}0; 4{,}4\rangle$	10	0,048	
9	$(4{,}4; 4{,}9\rangle$	10	0,060	
10	$(4{,}9; 5{,}8\rangle$	12	0,106	
11	$(5{,}8; 6{,}8\rangle$	10	0,106	
12	$(6{,}8; 8{,}7\rangle$	12	0,145	
13	$(8{,}7; \infty)$	11	0,107	
Celkem	-	132	1,000	-

χ^2 test dobré shody s očekávaným rozdělením

Řešení:

- Určíme očekávané četnosti.

i	Třídící interval (s)	Empirické četnosti O_i	Očekávané pravd. π_{0_i}	Očekávané četnosti E_i
1	$(-\infty; 1{,}5\rangle$	11	0,174	22,9
2	$(1{,}5; 1{,}8\rangle$	13	0,024	3,2
3	$(1{,}8; 2{,}0\rangle$	7	0,017	2,3
4	$(2{,}0; 2{,}5\rangle$	10	0,047	6,2
5	$(2{,}5; 2{,}9\rangle$	8	0,041	5,4
6	$(2{,}9; 3{,}6\rangle$	8	0,078	10,3
7	$(3{,}6; 4{,}0\rangle$	10	0,047	6,2
8	$(4{,}0; 4{,}4\rangle$	10	0,048	6,3
9	$(4{,}4; 4{,}9\rangle$	10	0,060	8,0
10	$(4{,}9; 5{,}8\rangle$	12	0,106	14,0
11	$(5{,}8; 6{,}8\rangle$	10	0,106	13,9
12	$(6{,}8; 8{,}7\rangle$	12	0,145	19,2
13	$(8{,}7; \infty)$	11	0,107	14,1
Celkem	-	132	1,000	-

$\leftarrow E_1 = 132 \cdot 0{,}174$

χ^2 test dobré shody s očekávaným rozdělením

Řešení:

- Určíme očekávané četnosti.

i	Třídící interval (s)	Empirické četnosti O_i	Očekávané pravd. π_{0_i}	Očekávané četnosti E_i
1	$(-\infty; 1{,}5\rangle$	11	0,174	22,9
2	$(1{,}5; 1{,}8\rangle$	13	0,024	3,2
3	$(1{,}8; 2{,}0\rangle$	7	0,017	2,3
4	$(2{,}0; 2{,}5\rangle$	10	0,047	6,2
5	$(2{,}5; 2{,}9\rangle$	8	0,041	5,4
6	$(2{,}9; 3{,}6\rangle$	8	0,078	10,3
7	$(3{,}6; 4{,}0\rangle$	10	0,047	6,2
8	$(4{,}0; 4{,}4\rangle$	10	0,048	6,3
9	$(4{,}4; 4{,}9\rangle$	10	0,060	8,0
10	$(4{,}9; 5{,}8\rangle$	12	0,106	14,0
11	$(5{,}8; 6{,}8\rangle$	10	0,106	13,9
12	$(6{,}8; 8{,}7\rangle$	12	0,145	19,2
13	$(8{,}7; \infty)$	11	0,107	14,1
Celkem	-	132	1,000	-

$\leftarrow E_{13} = 132 \cdot 0{,}107$

χ^2 test dobré shody s očekávaným rozdělením

Řešení:

- Ověříme předpoklady testu.

i	Třídící interval (s)	Empirické četnosti O_i	Očekávané pravd. π_{0_i}	Očekávané četnosti E_i
1	$(-\infty; 1{,}5\rangle$	11	0,174	22,9
2	$(1{,}5; 1{,}8\rangle$	13	0,024	3,2
3	$(1{,}8; 2{,}0\rangle$	7	0,017	2,3
4	$(2{,}0; 2{,}5\rangle$	10	0,047	6,2
5	$(2{,}5; 2{,}9\rangle$	8	0,041	5,4
6	$(2{,}9; 3{,}6\rangle$	8	0,078	10,3
7	$(3{,}6; 4{,}0\rangle$	10	0,047	6,2
8	$(4{,}0; 4{,}4\rangle$	10	0,048	6,3
9	$(4{,}4; 4{,}9\rangle$	10	0,060	8,0
10	$(4{,}9; 5{,}8\rangle$	12	0,106	14,0
11	$(5{,}8; 6{,}8\rangle$	10	0,106	13,9
12	$(6{,}8; 8{,}7\rangle$	12	0,145	19,2
13	$(8{,}7; \infty)$	11	0,107	14,1
Celkem	-	132	1,000	-

χ^2 test dobré shody s očekávaným rozdělením

Řešení:

- Ověříme předpoklady testu.

i	Třídící interval (s)	Empirické četnosti O_i	Očekávané pravd. π_{0_i}	Očekávané četnosti E_i
1	$(-\infty; 1{,}5\rangle$	11	0,174	22,9
2	$(1{,}5; 1{,}8\rangle$	13	0,024	3,2
3	$(1{,}8; 2{,}0\rangle$	7	0,017	2,3
4	$(2{,}0; 2{,}5\rangle$	10	0,047	6,2
5	$(2{,}5; 2{,}9\rangle$	8	0,041	5,4
6	$(2{,}9; 3{,}6\rangle$	8	0,078	10,3
7	$(3{,}6; 4{,}0\rangle$	10	0,047	6,2
8	$(4{,}0; 4{,}4\rangle$	10	0,048	6,3
9	$(4{,}4; 4{,}9\rangle$	10	0,060	8,0
10	$(4{,}9; 5{,}8\rangle$	12	0,106	14,0
11	$(5{,}8; 6{,}8\rangle$	10	0,106	13,9
12	$(6{,}8; 8{,}7\rangle$	12	0,145	19,2
13	$(8{,}7; \infty)$	11	0,107	14,1
Celkem	-	132	1,000	-

Všechny očekávané četnosti jsou větší nebo rovny 2, sloučíme-li třídy 2 a 3, bude splněn i silnější předpoklad – očekávané četnosti budou větší než 5.

χ^2 test dobré shody s očekávaným rozdělením

Řešení:

- Určíme pozorovanou hodnotu x_{OBS}.

i	Třídící interval (s)	Empirické četnosti O_i	Očekávané pravd. π_{0_i}	Očekávané četnosti E_i
1	$(-\infty; 1{,}5\rangle$	11	0,174	22,9
2	$(1{,}5; 2{,}0\rangle$	20	0,041	5,5
3	$(2{,}0; 2{,}5\rangle$	10	0,047	6,2
4	$(2{,}5; 2{,}9\rangle$	8	0,041	5,4
5	$(2{,}9; 3{,}6\rangle$	8	0,078	10,3
6	$(3{,}6; 4{,}0\rangle$	10	0,047	6,2
7	$(4{,}0; 4{,}4\rangle$	10	0,048	6,3
8	$(4{,}4; 4{,}9\rangle$	10	0,060	8,0
9	$(4{,}9; 5{,}8\rangle$	12	0,106	14,0
10	$(5{,}8; 6{,}8\rangle$	10	0,106	13,9
11	$(6{,}8; 8{,}7\rangle$	12	0,145	19,2
12	$(8{,}7; \infty)$	11	0,107	14,1
Celkem	-	132	1,000	-

$$x_{OBS} = \sum_{i=1}^{12} \frac{(O_i - E_i)^2}{E_i} = \frac{(11-22{,}9)^2}{22{,}9} + \cdots + \frac{(11-14{,}1)^2}{14{,}1} = 59{,}7$$

χ^2 test dobré shody s očekávaným rozdělením

Řešení:

- Určíme $p - hodnotu$.

i	Třídící interval (s)	Empirické četnosti O_i	Očekávané pravd. π_{0_i}	Očekávané četnosti E_i
1	$(-\infty; 1{,}5\rangle$	11	0,174	22,9
2	$(1{,}5; 2{,}0\rangle$	20	0,041	5,5
3	$(2{,}0; 2{,}5\rangle$	10	0,047	6,2
4	$(2{,}5; 2{,}9\rangle$	8	0,041	5,4
5	$(2{,}9; 3{,}6\rangle$	8	0,078	10,3
6	$(3{,}6; 4{,}0\rangle$	10	0,047	6,2
7	$(4{,}0; 4{,}4\rangle$	10	0,048	6,3
8	$(4{,}4; 4{,}9\rangle$	10	0,060	8,0
9	$(4{,}9; 5{,}8\rangle$	12	0,106	14,0
10	$(5{,}8; 6{,}8\rangle$	10	0,106	13,9
11	$(6{,}8; 8{,}7\rangle$	12	0,145	19,2
12	$(8{,}7; \infty)$	11	0,107	14,1
Celkem	-	132	1,000	-

$p\text{–}hodnota = 1 - F_0(59{,}7) = P(X > 59{,}7) \ll 0{,}001$, kde $X \sim \chi^2(k-1-h)$ je distribuční funkce χ^2 rozdělení s 9 $(= k - 1 - h = 12 - 1 - 2)$ stupni volnosti.

χ^2 test dobré shody s očekávaným rozdělením

Řešení:

X … časový odstup mezi průjezdy jednotlivých vozidel.

H_0: Časové odstupy mezi průjezdy jednotlivých vozidel mají norm. rozdělení.

H_A: Časové odstupy mezi průjezdy jednotlivých vozidel nemají norm. rozdělení.

$p\text{–}hodnota \ll 0{,}001$

Na hladině významnosti 0,05 zamítáme nulovou hypotézu, tj. nelze předpokládat, že časové odstupy mezi průjezdy jednotlivých vozidel mají normální rozdělení (χ^2 test dobré shody, $x_{OBS} = 59{,}7, df = 9, p - hodnota \ll 0{,}001$).

Kolmogorovovův – Smirnovův test

H_0: Náhodný výběr pochází z rozdělení se spojitou distribuční funkcí $F_0(x)$.

H_A: Náhodný výběr nepochází z rozdělení se spojitou distribuční funkcí $F_0(x)$.

POZOR!!!

Distribuční funkcí $F_0(x)$ musí být úplně specifikována.

Poznámky:

- Tímto testem nelze ověřit např. hypotézu, že výběr pochází z normálního rozdělení.
- To umožňuje až modifikovaný Kolmogorovův-Smirnovův test známý též pod názvem **Lillieforsův test**.

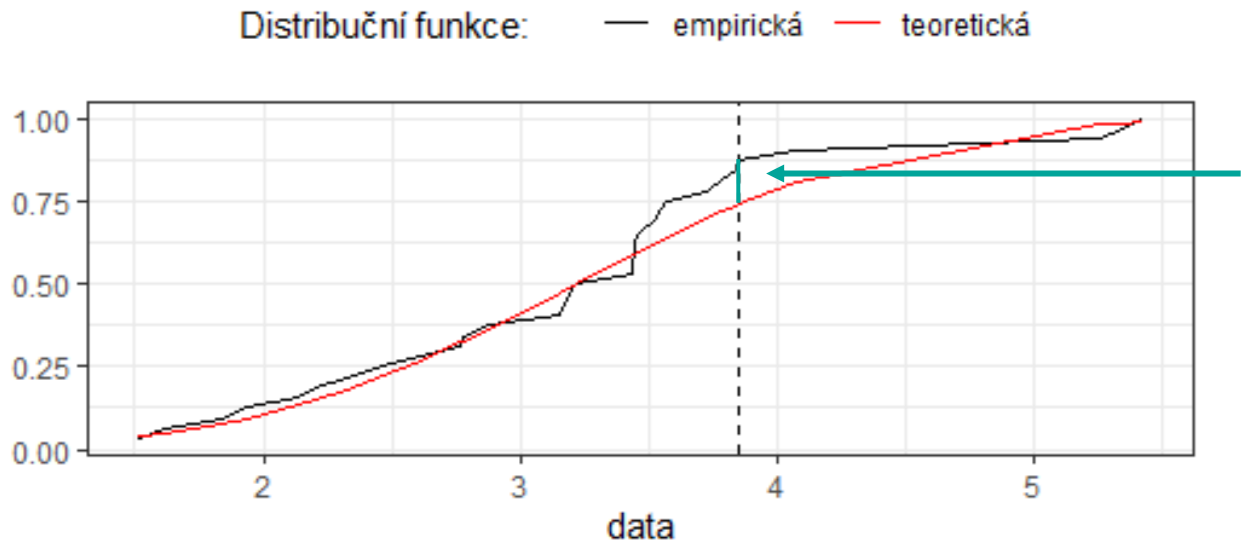
Kolmogorovovův – Smirnovův test

H_0: Náhodný výběr pochází z rozdělení se spojitou distribuční funkcí $F_0(x)$.

H_A: Náhodný výběr nepochází z rozdělení se spojitou distribuční funkcí $F_0(x)$.

- **Testové kritérium**

$$D_n = \sup|F_n(x) - F_0(x)| = max(D_1{}^*, D_2{}^*, \dots, D_n{}^*), \text{ kde } D_i{}^* = max\left\{\left|F_0(x) - \frac{i-1}{n}\right|; \left|F_0(x) - \frac{i}{n}\right|\right\} \text{ pro } i = 1, 2, \dots, n$$



max. odchylka teoretické a empirické distribuční funkce

Kolmogorovovův – Smirnovův test

H_0: Náhodný výběr pochází z rozdělení se spojitou distribuční funkcí $F_0(x)$.

H_A: Náhodný výběr nepochází z rozdělení se spojitou distribuční funkcí $F_0(x)$.

- **Testové kritérium**

$$D_n = \sup|F_n(x) - F_0(x)| = max(D_1{}^*, D_2{}^*, \dots, D_n{}^*), \text{ kde } D_i{}^* = max\left\{\left|F_0(x) - \frac{i-1}{n}\right|; \left|F_0(x) - \frac{i}{n}\right|\right\} \text{ pro } i = 1, 2, \dots, n$$

- **Rozhodnutí**
  - ✓ Nulovou hypotézu zamítáme, pokud $x_{OBS} > D_{n(\alpha)}$.
  - ✓ Je-li n malé $(n \leq 100)$, lze kritické hodnoty $D_{n(\alpha)}$ najít např. v tabulce T14.
  - ✓ Při velkých hodnotách n se kritické hodnoty $D_{n(\alpha)}$ aproximují pomocí vztahu $D_{n(\alpha)} \cong \sqrt{\frac{1}{2n} ln \frac{2}{\alpha}}$.

Ověření normality

- **Empirické (orientační – viz přednáška Explorační analýza)**
  - ✓ vizuální posouzení krabicového grafu (očekávaná symetrie), histogramu (tvar obálky histogramu odpovídá Gaussově křivce), Q-Q graf,
  - ✓ výběrová šikmost a špičatost je mezi -2 a 2,
  - ✓ srovnání průměru a mediánu (neměly by se příliš lišit)

- **Exaktní** (nejpoužívanější testy)
  - ✓ Shapirův-Wilkův test (původní test vhodný pro $n < 50$, dnes se používá Roystonova modifikace vhodná pro $n < 5000$ – v R: *shapiro.test()*)
  - ✓ Andersonův-Darlingův test
  - ✓ Kolmogorovův-Smirnovův test (Lillieforsova modifikace pro neúplně specifikovaný test)

Děkuji za pozornost!

martina.litschmannova@vsb.cz

VŠB TECHNICKÁ UNIVERZITA OSTRAVA | FAKULTA ELEKTROTECHNIKY A INFORMATIKY | KATEDRA APLIKOVANÉ MATEMATIKY