

Katedra pravděpodobnosti a matematické statistiky



MATEMATICKO-FYZIKÁLNÍ  
FAKULTA  
Univerzita Karlova

---

Kristína Mečiarová

**Q-Q grafy**  
Oborový seminár

---

2. novembra 2021

- ① Motivácia
- ② Normálne Q-Q grafy + bonus (kvíz)
- ③ Všeobecné Q-Q grafy
- ④ Q-Q grafy pre porovnanie 2 náhodných výberov
- ⑤ Využitie v lineárnej regresii

# Pripomenutie a značenie

## Q-Q grafy

- $\mathbf{X}_{(\cdot)} = (X_{(1)}, \dots, X_{(n)})^T$  ... usporiadaný náhodný výber
- $F_X(x) = P(X \leq x)$  ... distribučná funkcia n.v.  $X$
- $F_X^{-1}(u) = \inf\{x : F_X(x) \geq u\}$  ... kvantilová funkcia n.v.  $X$
- $q_X(\alpha) = F_X^{-1}(\alpha)$  ...  $\alpha$ -kvantil n.v.  $X$
- $u_\alpha = \Phi^{-1}(\alpha)$  ...  $\alpha$ -kvantil n.v.  $Y \sim N(0, 1)$

# Pripomenutie a značenie empirické odhady

## Q-Q grafy

- $\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}\{X_i \leq x\}$  ... empirická distribučná funkcia
- $\hat{F}_n^{-1}(\alpha) = \inf\{x : \hat{F}_n(x) \geq \alpha\}$  ... empirická kvantilová funkcia
- $\hat{u}_n(\alpha) = X_{(k_\alpha)}$  ... výberový  $\alpha$ -kvantil
  - $k_\alpha = \begin{cases} n\alpha, & n\alpha \in \mathbb{N}, \\ \lfloor n\alpha \rfloor + 1, & n\alpha \notin \mathbb{N}. \end{cases}$

# Motivačný príklad

## Q-Q grafy

- $X_1, \dots, X_{266} \dots$  náhodný výber z rozdelenia  $F_X$ , systolický tlak fajčiarov
- $Y_1, \dots, Y_{234} \dots$  náhodný výber z rozdelenia  $F_Y$ , systolický tlak nefajčiarov
- je niektorá skupina homogénnejšia?  $\Rightarrow F$ -test zhody rozptylov, predpoklad normality nutný!
- je priemerný tlak fajčiarov vyšší ako nefajčiarov? (normalita + zhoda rozptylov)  $\Rightarrow$  presný dvojvýberový  $t$ -test

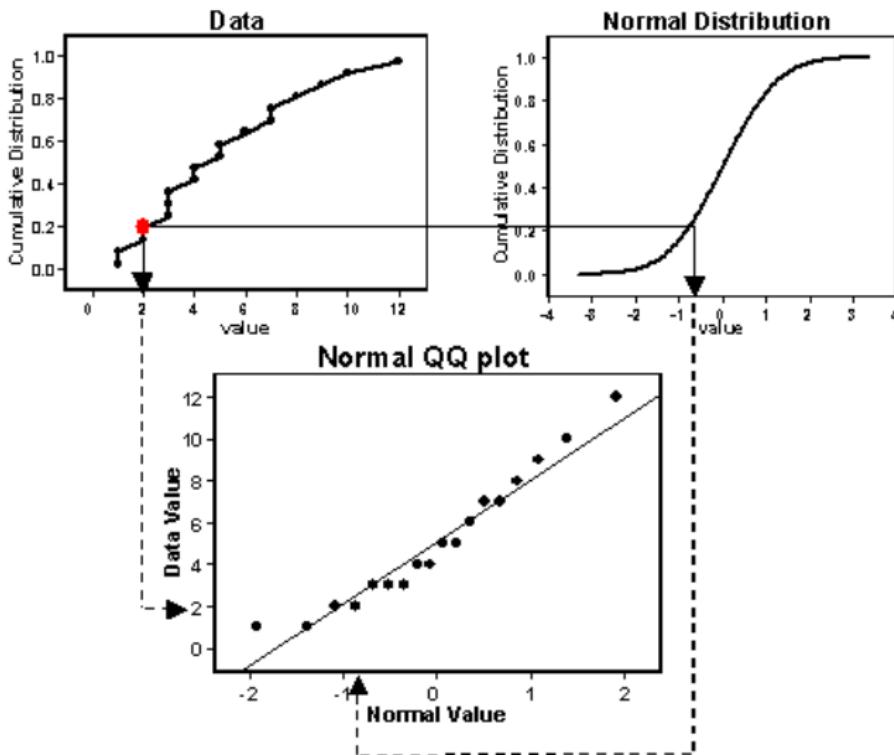
# Normálne Q-Q grafy

## Q-Q grafy

- zrovnanie výberových kvantilov náhodného výberu s teoretickými kvantilmi rozdelenia  $N(0, 1)$
- požadujeme lineárnu závislosť výberových a teoretických kvantilov
  - os  $x$  ... teoretické kvantily
  - os  $y$  ... výberové kvantily
- predpoklad  $X \sim N(\mu, \sigma^2)$ ,  $\mu \in \mathbb{R}$ ,  $\sigma^2 \in (0, \infty)$
- platí  $q_X(\alpha) = \mu + \sigma u_\alpha$  ?

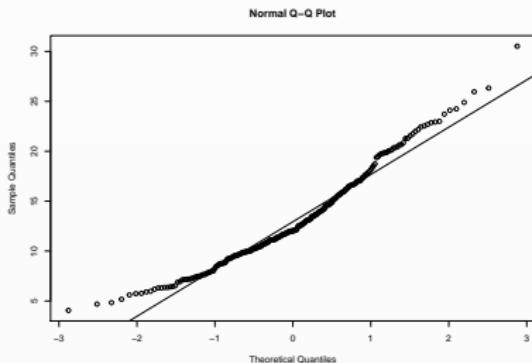
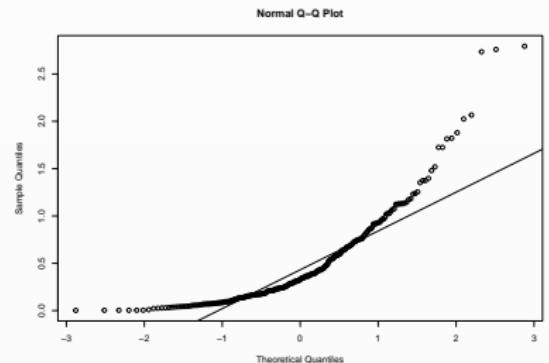
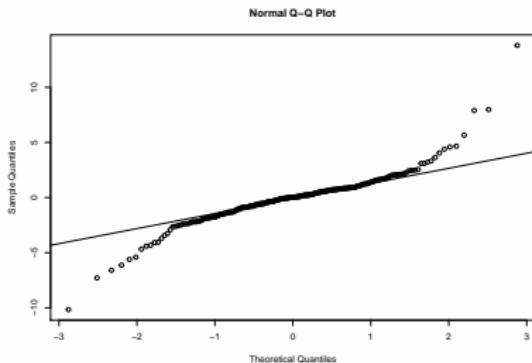
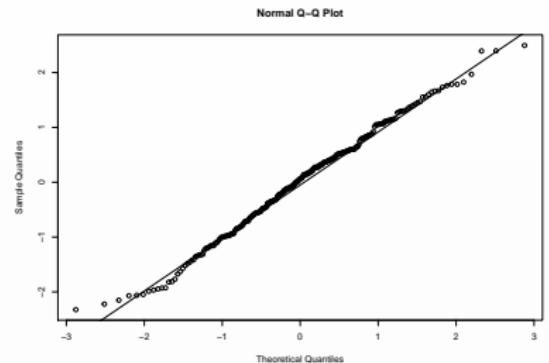
# Normálne Q-Q grafy

## Q-Q grafy



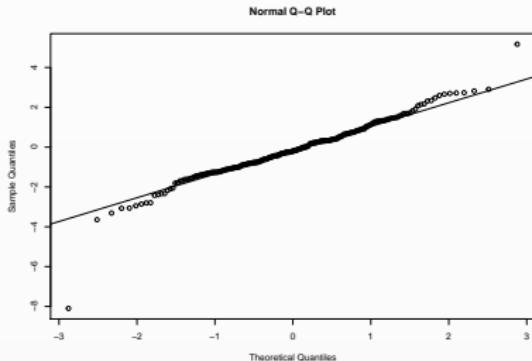
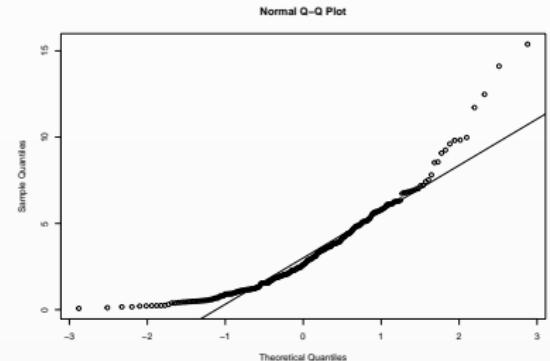
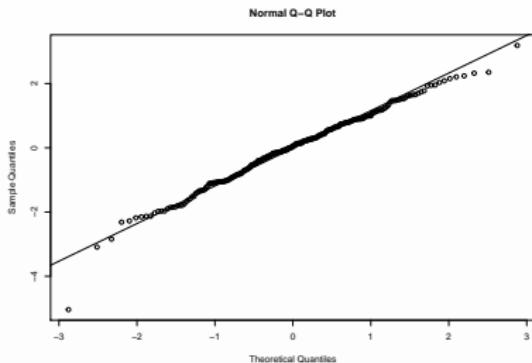
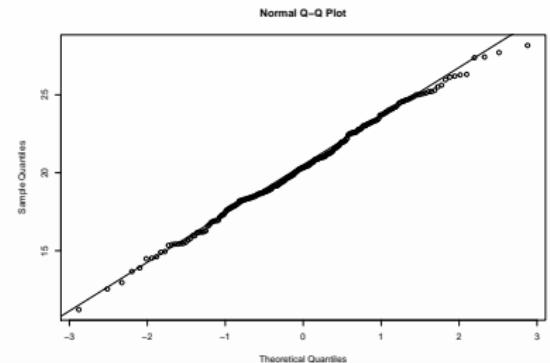
# Kvíz: normálne Q-Q grafy pre rôzne rozdelenia

## Q-Q grafy



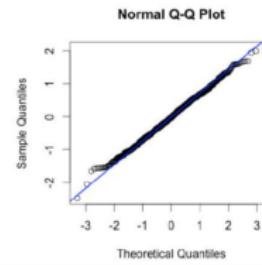
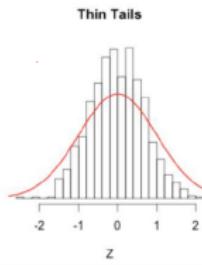
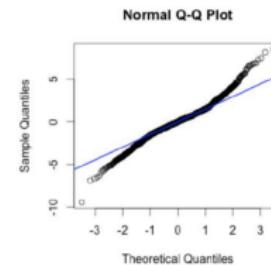
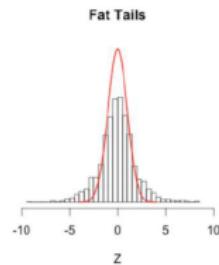
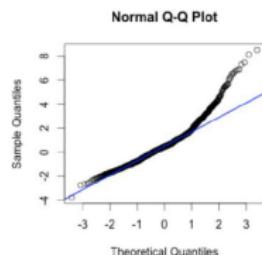
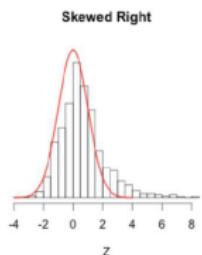
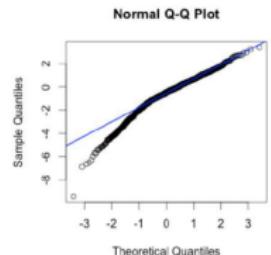
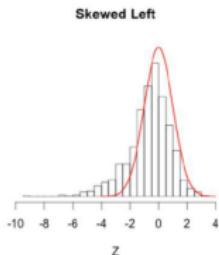
# Kvíz: normálne Q-Q grafy pre rôzne rozdelenia

## Q-Q grafy



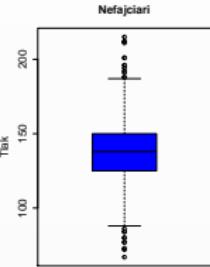
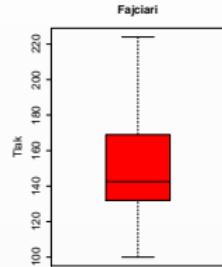
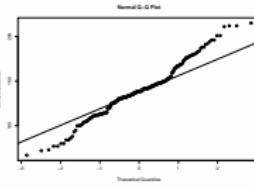
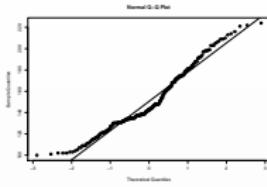
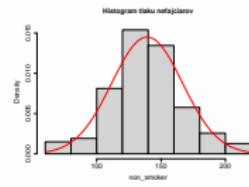
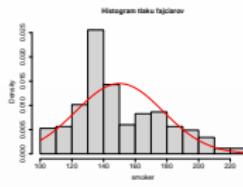
# Q-Q graf vs. histogram

## Q-Q grafy



# Motivačný príklad grafy

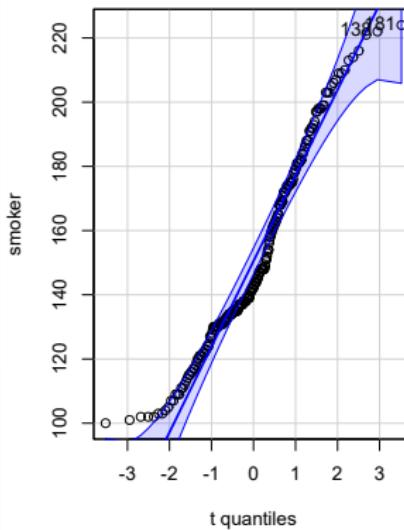
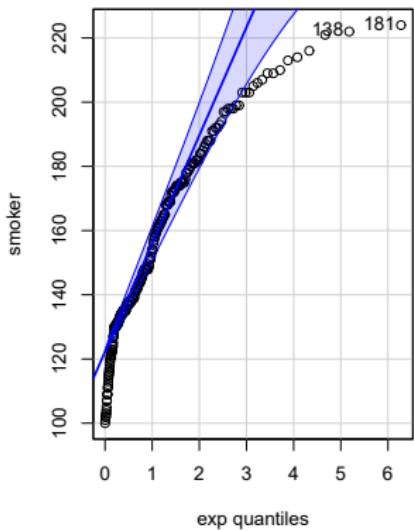
## Q-Q grafy



# Všeobecný Q-Q graf

## Q-Q grafy

- zrovnanie výberových kvantilov s kvantilmami nejakého predom zvoleného rozdelenia



### Normálny Q-Q graf

```
> qnorm(smoker)
> qline(smoker)
```

### Všeobecný Q-Q graf

```
> qqPlot(smoker, distribution="exp")
> qqPlot(smoker, distribution="t", df=13)
> y <- rchisq(500, df = 3)
> qqplot(qchisq(ppoints(500), df = 3), y)
> qpline(y, distribution = function(p) qchisq(p, df = 3), probs = c(0.1, 0.6))
```

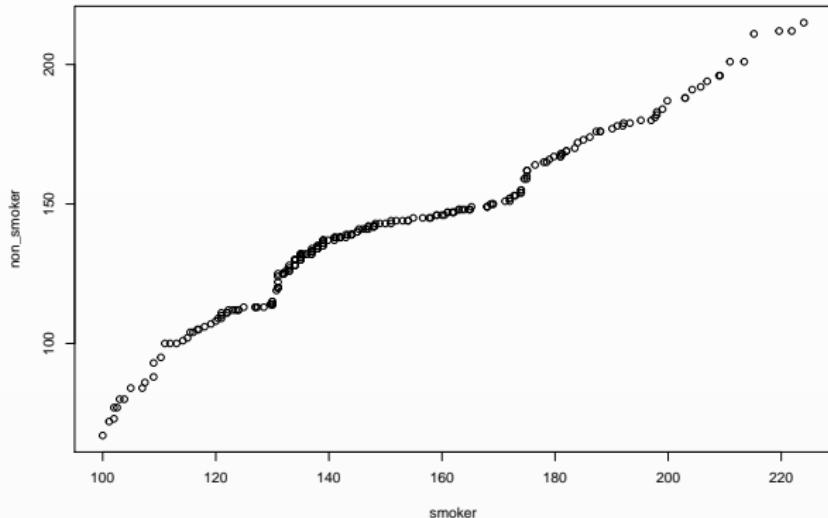
### Q-Q graf pre porovnanie 2 náhodných výberov

```
> qqplot(smoker, non_smoker)
```

# Motivačný príklad inak

## Q-Q grafy

- sú rozdelenia tlaku fajčiarov a nefajčiarov zhodné?



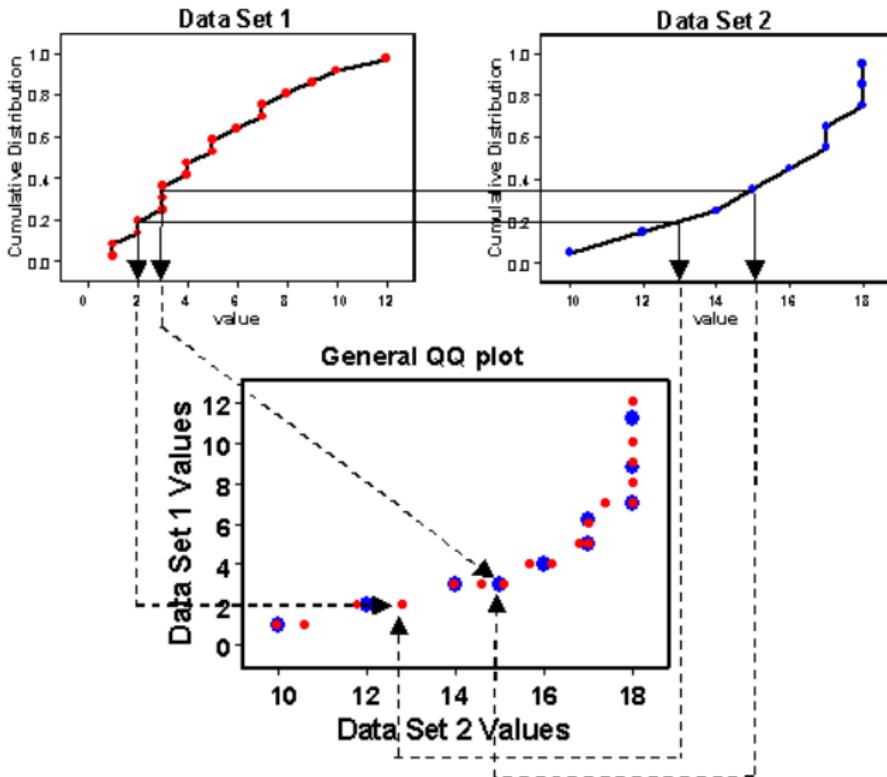
# Q-Q graf pre zrovnanie 2 výberov

## Q-Q grafy

- zrovnanie výberových kvantilov 2 náhodných výberov
- $X \sim F_X, Y \sim F_Y$
- oba výbery pochádzajú z rovnakého rozdelenia  $\implies$  výberové kvantily ležia "približne" na priamke  $y = x$

# Q-Q graf pre zrovnanie 2 výberov

## Q-Q grafy



# Použitie v lineárnej regresii

## Q-Q grafy

- lineárny model pre všeobecné dáta  $\mathbf{Y}|\mathbb{X} \sim (\mathbb{X}\beta, \sigma^2\mathbf{I}_n)$ ,  
 $\beta = (\beta_0, \dots, \beta_{k-1}) \in \mathbb{R}^k$  a  $\sigma^2 < \infty$  neznáme parametre
- $\epsilon = (\epsilon_1, \dots, \epsilon_n) = \mathbf{Y} - \mathbb{X}\beta$  ... chybový vektor
- predpoklad normality (A4)  $Y_i|\mathbf{X}_i = \mathbf{x} \sim N(\mathbf{x}^T \beta, \sigma^2)$ ,  
 $\epsilon_i|\mathbf{X}_i = \mathbf{x} \sim N(0, \sigma^2)$ ,  $i = 1, \dots, n$
- $\mathbf{U} = \mathbf{Y} - \hat{\mathbf{Y}} = \mathbb{M}\mathbf{Y} = \mathbb{M}\epsilon$ , kontrolujeme  $\mathbf{U}|\mathbb{X} \sim N_n(\mathbf{0}_n, \sigma^2\mathbb{M})$

# Použitie v lineárnej regresii

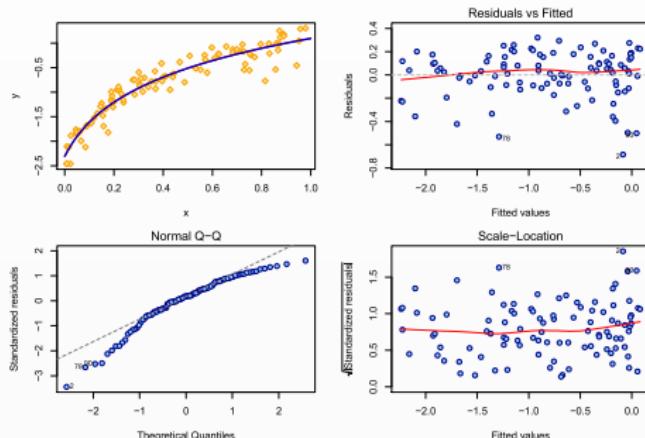
## Q-Q grafy

### Illustrations

#### Nonnormal errors

**True:**  $Y = \log(0.1 + x) + \varepsilon, \quad \varepsilon \sim \text{Gumbel}.$

**Model:**  $Y = \beta_0 + \beta_1 \log(0.1 + x) + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2).$



# Použitie v lineárnej regresii

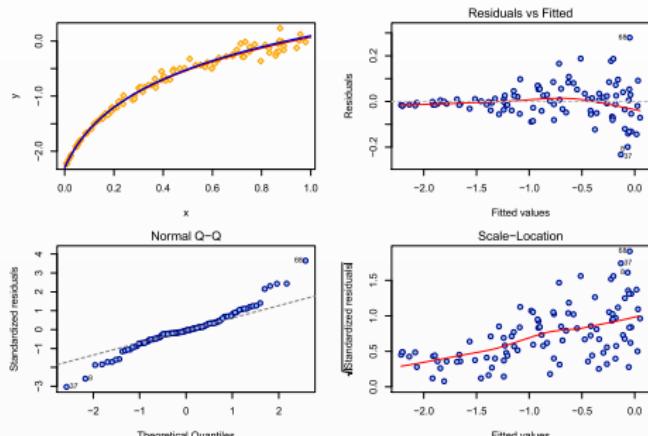
## Q-Q grafy

### Illustrations

#### Heteroscedasticity

**True:**  $Y = \log(0.1 + x) + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, (0.2x)^2)$ .

**Model:**  $Y = \beta_0 + \beta_1 \log(0.1 + x) + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2)$ .



Ďakujem za pozornosť!