

Databázové systémy a SQL

A decorative background graphic consisting of a series of vertical, stylized elements that resemble a bar chart or a sequence of data points. Each element has a central circle and a vertical line extending above and below it, with a horizontal line connecting them. The elements are arranged in a slightly curved line across the top right of the slide.

Lekce 11

Daniel Klimeš

- COVAR_POP
- COVAR_SAMP
- CORR
 - CORR (Pearson's correlation coefficient)
 - CORR_S (Spearman's coefficient)
 - CORR_K (Kendall's coefficient)
- **Lineární regrese**
 - REGR_INTERCEPT (závislá, nezávislá)
 - REGR_SLOPE (závislá, nezávislá)
- **Statistické testy**
 - STATS_BINOMIAL_TEST - binomický test
 - STATS_CROSSTAB – kontingenční tabulka
 - STATS_F_TEST
 - STATS_KS_TEST - Kolmogorov-Smirnov test
 - STATS_MODE - modus
 - STATS_MW_TEST - Mann Whitney test
 - STATS_ONE_WAY_ANOVA - ANOVA
 - STATS_T_TEST_*
 - STATS_WSR_TEST - Wilcoxon Signed Ranks test

```
SELECT COVAR_POP(hmotnost .value, vyska.value) kovariance_pop,
COVAR_SAMP(hmotnost.value, vyska.value) kovariance_sample,
CORR(hmotnost.value, vyska.value) pearson,
CORR_S(hmotnost.value, vyska.value) spearman
FROM eav_real hmotnost, eav_real vyska
WHERE hmotnost.question_id = 6081 AND vyska.question_id = 6083
AND hmotnost.subheader_id = vyska.subheader_id
```

59,276

59,588

0,4799

0,4952

```
INSERT INTO linear (x,y)
SELECT ROWNUM, 3*ROWNUM +5
FROM questions WHERE ROWNUM < 50
```

```
SELECT REGR_SLOPE (y, x),
REGR_INTERCEPT (y, x) FROM linear
```

```
SELECT REGR_SLOPE (hmotnost.value, vyska.value),
REGR_INTERCEPT (hmotnost.value, vyska.value)
FROM eav_real hmotnost, eav_real vyska
WHERE hmotnost.question_id = 6081 AND vyska.question_id = 6083
AND hmotnost.subheader_id = vyska.subheader_id
```

STATISTIC	The observed value of t
DF	Degree of freedom
ONE_SIDED_SIG	One-tailed significance of t
TWO_SIDED_SIG	Two-tailed significance of t

```

SELECT AVG(hmotnost.value) prumer,
STATS_T_TEST_ONE (hmotnost.value, 72, 'TWO_SIDED_SIG')
two_side,
STATS_T_TEST_ONE (hmotnost.value, 72, 'ONE_SIDED_SIG')
one_side,
STATS_T_TEST_ONE (hmotnost.value, 72, 'DF') df
FROM eav_real hmotnost, eav_real vyska
WHERE hmotnost.question_id = 6081 AND vyska.question_id = 6083

```

```

SELECT
STATS_T_TEST_INDEP (mod (rownum,2), value, 'TWO_SIDED_SIG')
two_side,
STATS_T_TEST_INDEP (mod (rownum,2), value, 'DF') df,
STATS_T_TEST_INDEPU (mod (rownum,2), value, 'TWO_SIDED_SIG')
two_side_u,
STATS_T_TEST_INDEPU (mod (rownum,2), value, 'DF') df_u
FROM
(SELECT vyska.value
FROM eav_real hmotnost, eav_real vyska
WHERE hmotnost.question_id = 6081 AND vyska.question_id = 6083
AND hmotnost.subheader_id = vyska.subheader_id
order by vyska.subheader_id)

```

- Zjistěte průměrnou hodnotu, minimum , maximum otázky XY
- Vypište věk pacientů při zařazení do studie XY, zvláště muži, ženy
- Vypište hodnotu otázky XY pro všechny pacienty (včetně nevyplněných)
- Jaký je průměrný počet vyplněných formulářů na pacienta u studie XY
- Vypište seznam otázek, které nebyly dosud nikdy vyplněny
- Vypište počty žen a mužů zařazených v jednotlivých studiích

- Skupina otázek XY
 - Kolik má otázek, jaké datové typy, u číselníků jaké jsou možnosti
 - Jaká je vyplněnost, N, minimum, maximum, průměr, SD, maximální délka textu
 - Váš vlastní názor k vyplněnosti
 - Nějaká závislost mezi vyplněním otázek?

- Přehled max 2 strany A4
- Příloha – Způsob vytažení dat – komentované SQL, pracovní postup