

# Databázové systémy a SQL

## Lekce 7 - Statistické funkce

Daniel Klimeš

- COVAR\_POP – Kovariance populační
- COVAR\_SAMP – Kovariance vzorková
- CORR - Pearsonův korelační koeficient
- CORR\_S - Spearmanův koeficient
- CORR\_K - Kendallův koeficient)

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

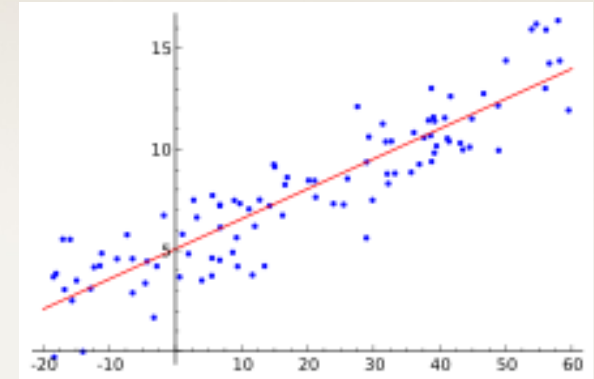
0,4952

$$Y = \beta x + \alpha$$

```
CREATE table regr_test as
SELECT ROWNUM x, 3*ROWNUM +5 y
FROM questions WHERE ROWNUM < 50
```

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

```
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
```



- Jednovzorkové (one sample)
- Párové uspořádání
- Nepárové (nezávislé) uspořádání
  
- Výstupní hodnoty testu

STATISTIC	Výsledek testu
DF	Stupeň volnosti
ONE_SIDED_SIG	Jednostranná významnost
TWO_SIDED_SIG	Oboustranná významnost

```

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

```

- Parametr měřený před a po zásahu (léčba)

```

SELECT
  STATS_T_TEST_PAired (pred.value, po.value, 'TWO_SIDED_SIG'),
  AVG(pred.value), AVG(po.value)
FROM eav_real pred, eav_real po
WHERE pred.question_id = 917 AND po.question_id = 918
  AND pred.subheader_id = po.subheader_id
  AND ROWNUM <= 100
  
```

- Data měřená na dvou nezávislých vzorcích

```

SELECT
STATS_T_TEST_INDEP (sex, sysdate – date_of_birth, 'TWO_SIDED_SIG')
two_side,
STATS_T_TEST_INDEP (sex, sysdate – date_of_birth 'DF') df,
STATS_T_TEST_INDEPU (sex, sysdate – date_of_birth 'TWO_SIDED_SIG')
two_side_u,
STATS_T_TEST_INDEPU (sex, sysdate – date_of_birth 'DF') df_u
FROM patients
WHERE sex IN ('F','M') and date_of_birth IS NOT NULL
    
```

- porovnání více než 2 vzorků
- analyzujeme účinek jednoho faktoru na zkoumanou proměnnou

```

SELECT
  STATS_ONE_WAY_ANOVA(riziko.value, delka_hosp.value, 'F_RATIO') f_ratio,
  STATS_ONE_WAY_ANOVA(riziko.value, delka_hosp.value, 'SIG') p_value

FROM eav_real delak_hosp, eav_int riziko,
     event_subheader es_hosp, event_subheader es_riz

WHERE delka_hosp.question_id = 3482
AND es_hosp.subheader_id = delka_hosp.subheader_id
AND es_riz.subheader_id = riziko.subheader_id
AND riziko.question_id = 3442 and es_riz.header_uid = es_hosp.header_uid

```

- 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