07/29 DM 下午場

洪子軒

Sent: Friday, July 29, 2016 2:18 PM
To: 洪子軒

【其他】

http://hamelg.blogspot.tw/2015/11/python-for-data-analysis-part-24.html

t-test

http://hamelg.blogspot.tw/2015/11/python-for-data-analysis-part-16_23.html

The t-test works well when dealing with two groups, but sometimes we want to compare more than two groups at the same time. For example, if we wanted to test whether voter age differs based on some categorical variable like race, we have to compare the means of each level or group the variable. We could carry out a separate t-test for each pair of groups, but when you conduct many tests you increase the chances of false positives. The analysis of variance or ANOVA is a statistical inference test that lets you compare multiple groups at the same time.

```
from statsmodels.stats.multicomp import pairwise_tukeyhsd

tukey = pairwise_tukeyhsd(endog=voter_age,     # Data
                          groups=voter_race,   # Groups
                          alpha=0.05)          # Significance level

tukey.plot_simultaneous()    # Plot group confidence intervals
plt.vlines(x=49.57,ymin=-0.5,ymax=4.5, color="red")

tukey.summary()              # See test summary
```

--
洪子軒 Tzu-Hsuan Hung
中華電信研究院 巨量資料所
TEL: (03)-4245128
Email: Lucas@cht.com.tw
32661桃園市楊梅區電研路99號
--