

24.12.04 (Hypothesis_testing_피드백)

- 60210021 박세민

(팀원 피드백) 날짜와 객실 유형은 호텔 유형 선택에 영향을 줄 것 같음. 포함에서 함께 돌려보면 유의미한 결과가 나오지 않을까?

=> arrival_date_month: 호텔에 도착한 월

arrival_date_week_number: 호텔에 도착한 주

arrival_date_day_of_month: 호텔에 도착한 일

reserved_room_type: 예약된 객실 유형(익명성을 위해 명칭 대신 코드로 표시)

위의 네 개의 컬럼을 독립변수 후보 컬럼으로 추가함.

(팀원 피드백) 정규 분포를 가정되지 못할 때는 Levene Test를 사용하는 if문을 만들어야 하지 않습니까? (Levene Test: 데이터의 분포에 덜 민감하며, 정규성 가정이 위배되더라도 비교적 안정적인 결과를 제공)

=> 정규성을 만족하지 못하면 바로 비모수 검정을 사용하면 되니까. 정규성을 만족했을 경우의 등분산성 검증만 처리했습니다. 이에 대한 적절한 주석을 추가하도록 하겠습니다.

```
from scipy.stats import kstest, bartlett, ttest_ind, ranksums

print("=== T-test 및 정규성/등분산성 검토 ===")
for var in numeric_cols:
    if df[var].dtype in ['int64', 'float64']:
        group1 = df[df['hotel'] == 'Resort Hotel'][var].dropna()
        group2 = df[df['hotel'] == 'City Hotel'][var].dropna()
        print(np.shape(group1), np.shape(group2))

        # 정규성 검증 (shapiro test-> KS test로 대체)
        stat_g1, p_g1 = kstest(group1, 'norm', args=(group1.mean(), group1.std()))
        stat_g2, p_g2 = kstest(group2, 'norm', args=(group2.mean(), group2.std()))
        is_normal = (p_g1 > 0.05) and (p_g2 > 0.05)
        print(f"{var} - 정규성 검증: Group1 p-value = {p_g1:.10f}, Group2 p-value = {p_g2:.10f}, Normal")

        # 등분산성 검증 (bartlett test)
        stat_var, p_var = bartlett(group1, group2)
        is_equal_var = p_var > 0.05
        print(f"{var} - 등분산성 검증: Bartlett p-value = {p_var:.10f}, Equal Variance = {is_equal_var}")

        # T-test 또는 대안적 방법 선택
        if is_normal:
            if is_equal_var: # 등분산 two sample t-test
                t_stat, p_value = ttest_ind(group1, group2, equal_var=True)
                test_type = "T-test (Equal Variance)"
            else: # 이분산 two sample t-test
                t_stat, p_value = ttest_ind(group1, group2, equal_var=False)
                test_type = "T-test (Unequal Variance)"
        else: # 비모수 검증
            t_stat, p_value = ranksums(group1, group2)
            test_type = "Rank sums Test (Non-Normal)"

        print(f"{var}: {test_type}, t-statistic = {t_stat:.4f}, p-value = {p_value:.10f}")
print("-" * 50)
```