
State and verify assumptions.

1) The datafile **SFO 2009.csv** shows passenger satisfaction at the SFO airport; this survey was taken in 2009. The columns E through K contain Likert scale 1-5 data with 1=Poor and 5 = Excellent. Any value in the columns E through K above 6 is a missing value and should be converted to NA before working on the following problems. (Hint: `D[X][D[X > 6] <- NA`)

(a) Draw boxplots of the variables RATE AIRPORT, RATE RESTAURANT, RATE RETAIL, and RATE SAFETY.

(b) Let $p.S$ = proportion of passengers who rate the SFO safety ≥ 4 . Compute a 95% confidence interval for $p.S$.

(c) Let $p.A$ = proportion of passengers who rate the SFO airport ≥ 4 . Compute a 95% confidence interval for $p.A$.

(d) From the confidence intervals above, can we conclude that $p.S = p.A$?

(e) Use testing hypotheses method to compare $p.S$ and $p.A$.

2) The datafile **Coffee Shop Spending.csv** shows amounts customers spend at two coffee shops A and B.

(a) Compute 95% confidence intervals for true mean spend amounts $\mu(A)$ and $\mu(B)$; are the two means $\mu(A)$ and $\mu(B)$ equal?

(b) Use testing hypotheses method to compare $\mu(A)$ and $\mu(B)$.

(c) Let p = proportion of customers spending more than \$12 at the Coffee Shop.

Compute 95% confidence intervals for the difference in proportions $p_A - p_B$.

Are p_A and p_B equal?