Comparison

COR1-GB.1305 – Statistics and Data Analysis

More p-values

1. Suppose we perform a hypothesis test and we observe a p-value of $p = .02$. True or false: There is a 2% chance that the null hypothesis is true.

2. Suppose we perform a hypothesis test and we observe a p-value of $p = .02$. True or false: If we reject the null hypothesis, then there is a 2% chance of making a type I error.

3. Suppose we perform a hypothesis test and we observe a T test statistic $t = -2.02$, corresponding to a p-value of $p = .02$. True or false: If we were to repeat the experiment and the null hypothesis were actually true, then there would be a 2% chance of observing a test statistic at least as extreme as $t = -2.02$.

Confidence Intervals for Comparing Means

4. Recall the class survey. Seventeen female and thirty male students filled out the survey, reporting (among other variables) their GMAT scores and interest levels in the course. We will use this data to compare females and males.

   (a) What are the relevant populations?

   (b) For the 14 female respondents who reported their GMAT scores, the mean was 721 and the standard deviation was 27. For the 28 male respondents, the mean was 720 and the standard deviation was 39. Find a 95% confidence interval for the difference in population means.

   (c) For the 17 female respondents who reported their interest levels in the course (1–10), the mean was 5.8 and the standard deviation was 1.8. For the 30 male respondents, the mean was 6.3 and the standard deviation was 2.1. Find a 95% confidence interval for the difference in population means.

   (d) For the confidence intervals you constructed in parts (b) and (c) to be valid, what assumptions need to be satisfied? How could you check these assumptions?

Hypothesis Tests for Comparing Means

5. Consider again the class survey data. We will use the data to evaluate whether or not there is a significant difference between the female and the male population means.

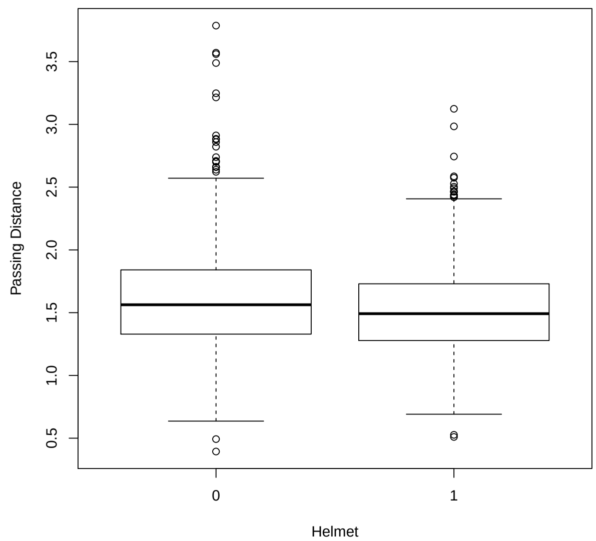
   (a) For the 14 female respondents who reported their GMAT scores, the mean was 721 and the standard deviation was 27. For the 28 male respondents, the mean was 720 and the standard deviation was 39. If the population means were equal what would be the chance of seeing a difference in sample means as large as observed?

   (b) For the 27 female respondents who reported their interest levels in the course (1–10), the mean was 5.8 and the standard deviation was 1.8. For the 30 male respondents, the mean was 6.3 and the standard deviation was 2.1. If the population means were equal what would be the chance of seeing a difference in sample means as large as observed?

   (c) What is the relationship between the confidence intervals in Question 4 and your answers to parts (a) and (b)?

Case Study: Bicycle Passing Distance

6. Here are boxplots of the passing distances (in meters) for a bike rider with and without a helmet. Is there evidence that the passing distance differs when the rider has a helmet?



Here are the sample statistics for the passing distance without a helmet: $n_1 = 1206$, $\bar{x}_1 = 1.61$, $s_1 = 0.405$. Here are the sample Here are the sample statistics for the passing distance with a helmet: $n_2 = 1149$, $\bar{x}_2 = 1.52$, $s_2 = 0.354$.

Formulate the problem as a hypothesis test, using significance level 5%.

(a) What are the populations?

(b) What are the null and alternative hypotheses?

(c) What are the samples?

(d) What is the test statistic?

(e) Approximately what is the p-value and the result of the test?

(f) Find a 95% confidence interval for the difference in passing difference with and without a helmet.