

Analytics Foundations: Problem Set 3

Today's dataset comes from a bike sharing company (Capital Bike Share). Each *hour*, the number of riders (**cnt**) is given, along with various other attributes as shown in the table below:

cnt	count of total rental bikes including both casual and registered
dteday	date
instant	record index (ID)
season	season (1:springer, 2:summer, 3:fall, 4:winter)
yr	year (0: 2011, 1:2012)
mnth	month (1 to 12)
hr	hour (0 to 23)
holiday	whether day is holiday or not
weekday	day of the week
workingday	if day is neither weekend nor holiday is 1, otherwise is 0.
weathersit	- 1: Clear, Few clouds, Partly cloudy, Partly cloudy - 2: Mist + Cloudy, Mist + Broken clouds, Mist + Few clouds, Mist - 3: Light Snow, Light Rain + Thunderstorm + Scattered clouds, Light Rain + Scattered clouds - 4: Heavy Rain + Ice Pallets + Thunderstorm + Mist, Snow + Fog
temp	Normalized temperature in Celsius. The values are divided to 41 (max)
atemp	Normalized feeling temperature in Celsius. The values are divided to 50 (max)
hum	Normalized humidity. The values are divided to 100 (max)
windspeed	Normalized wind speed. The values are divided to 67 (max)
casual	count of casual users
registered	count of registered users

Source: <http://archive.ics.uci.edu/ml/datasets/Bike+Sharing+Dataset#>

1. Does the weather situation (**weathersit** categorical) *appear* to have an effect on the number of riders (**cnt** continuous)?
 - a. What type of plot would help you decide?
 - b. Perform a statistical test to confirm your conclusion.

2. Repeat #1 to explore the effect of season (**season**) and then holidays (**holiday**).