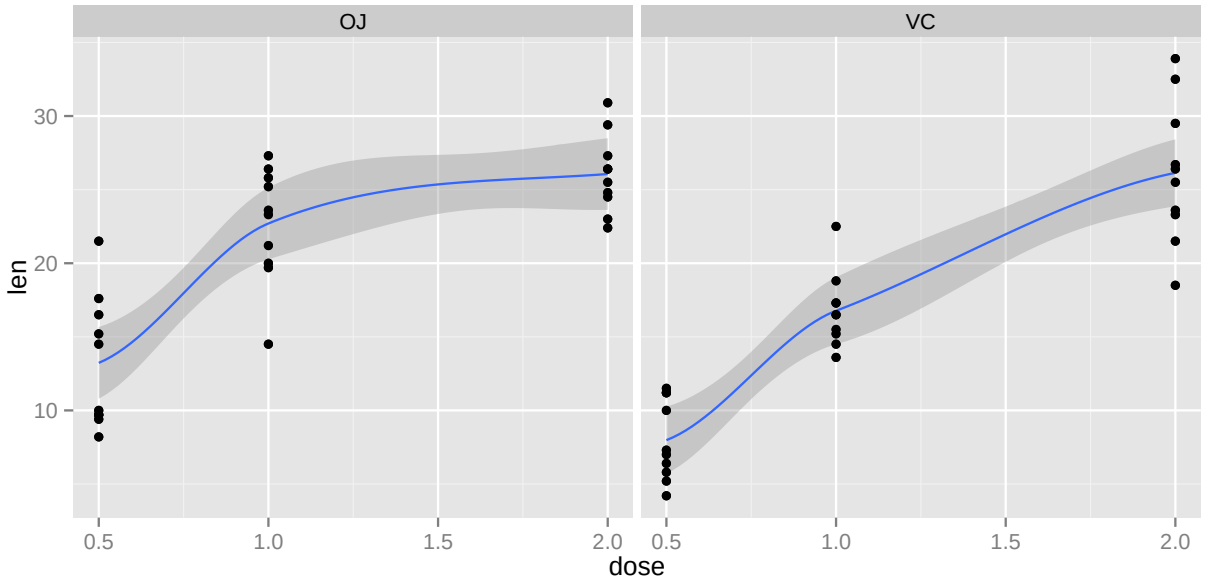
Statistical Inference - Course Project Part 2

Wednesday, August 20, 2014

This is the second part of the project for the Statistical Inference class. It consists of analyzing the ToothGrowth data in the R datasets package and executing the following tasks.

1. Load the ToothGrowth data and perform some basic exploratory data analyses.

   The ToothGrowth dataset contains the length of odontoblasts (teeth) in each of 10 guinea pigs at each of three dose levels of Vitamin C (0.5, 1, and 2 mg) with each of two delivery methods (orange juice - OJ or ascorbic acid - VC). The graph below illustrates the variation in tooth length, concerning each dose, for the two methods used.



2. Provide a basic summary of the data.

supp	dose	len.mean	len.std	len.median	len.min	len.max
OJ	0.5	13.23	4.460	12.25	8.2	21.5
OJ	1	22.70	3.911	23.45	14.5	27.3
OJ	2	26.06	2.655	25.95	22.4	30.9
VC	0.5	7.98	2.747	7.15	4.2	11.5
VC	1	16.77	2.515	16.50	13.6	22.5
VC	2	26.14	4.798	25.95	18.5	33.9

3. Use confidence intervals and hypothesis tests to compare tooth growth by supp and dose. (Use the techniques from class even if there's other approaches worth considering).

   For this task, let's consider that each individual belonging to a set of 10 guinea pigs grouped by "supp" and "dose" was chosen by randomization, so it's reasonable to just compare two groups with a t confidence interval or t test.

Based on above, the main goal is to compare the tooth length mean related to a dose of OJ with the tooth length mean related to the same dose of CV. Just looking at the previous graph and table, without any formal statistical test, it does appear that the average tooth length induced by OJ is greater than the average tooth length induced by CV, except for the 2 mg dose in which they appear to be the same. Let's actually look at it using a formal confidence interval of 95%. The 95th percentile of the t distribution with (10+10-2) 18 df is **1.734**.

Once the data documentation said nothing about how data was collected, whether paired or independent, it is reasonable to assume, from the data structure, that they are independent groups. However, for comparison purpose, both cases were considered.

Once it has been assumed randomness, it would be reasonable to expect that the variance is the same for each two groups (OJ and VC) of the same dose. However, when there is some doubt, like in the present case, a different variance per group must be assumed. For the pourpose of this analisys, both were considered.

Hypothesis 1 (dose of 0.5 mg): $H_0 : \mu_{OJ} = \mu_{CV}$ and $H_a : \mu_{OJ} > \mu_{CV}$

analysis	variance	t	low.limit	upper.limit
paired	equal	2.979	2.020	Inf
independent	equal	3.170	2.378	Inf
independent	different	3.170	2.346	Inf

Hypothesis 2 (dose of 1 mg): $H_0 : \mu_{OJ} = \mu_{CV}$ and $H_a : \mu_{OJ} > \mu_{CV}$

analysis	variance	t	low.limit	upper.limit
paired	equal	3.372	2.706	Inf
independent	equal	4.033	3.380	Inf
independent	different	4.033	3.356	Inf

Hypothesis 3 (dose of 2 mg): $H_0 : \mu_{OJ} = \mu_{CV}$ and $H_a : \mu_{OJ} > \mu_{CV}$

analysis	variance	t	low.limit	upper.limit
paired	equal	-0.0426	-3.523	Inf
independent	equal	-0.0461	-3.087	Inf
independent	different	-0.0461	-3.134	Inf

4. State your conclusions and the assumptions needed for your conclusions.

This analysis takes into account both the output of the t test and the confidence interval (CI). The test results of hypotheses 1 and 2 (0.5 mg and 1 mg) show that the t statistic is greater than the 95th percentile (1.734) for all the situations tested. Hence, H_0 must be rejected. The CIs are all entirely above zero which allows us to accept that $\mu_{OJ} - \mu_{CV} > 0$. Comparing equal and different variances for independent groups, it can be seen that CIs are wider when different variances are considered. Based on these two first hypotheses, we can suggest that the efficiency of OJ, concerning tooth growth, is higher than that of VC. For the hypothesis 3 (2 mg), on the other hand, the t statistic is lower than 1.734 and the CI contains zero, so we need to admit zero as a possibility for the population difference between the two groups. Hence, we fail to reject H_0 and we can infer that, for higher doses, OJ and CV are equally efficient in helping tooth growth.

P.S.: The markdown file used to generate this report can be found in https://github.com/Fpschwartz1/StatisticalInference_CourseProject/blob/master/CourseProject_Part2.Rmd