

Homework 2

Aji John

Winter Quarter 2018

This template is meant to be a helpful starting point for writing up homeworks for students who want to use Rmarkdown. Modify/distribute it as you see fit. At a minimum, you'll want to remove these comments when you turn in your assignments. For additional information about using Rmarkdown, see the help files and links posted to the course Canvas website. Remember, raw R code/output is not acceptable in your homework.

Question 1

In this question, you will perform statistical analyses evaluating an association between serum creatinine levels (measured from blood) and 5 year all-cause mortality by comparing mean creatinine levels across groups defined by vital status at 5 years using a t-test that presumes homoscedasticity (i.e., equal variances across groups). As this problem is directed toward illustrating correspondences between the t-test and linear regression, you do not need to provide full statistical inference for this problem. Instead, just answer the following questions.

(a)

What are the sample size, sample mean and sample standard deviation of serum creatinine levels among subjects who survived at least 5 years?

n	mean	sd
600	1.0325	0.244322

- Two samples skipped as 'NA' featured in them.
- Units for mean is mg/dl

(b)

What are the sample size, sample mean and sample standard deviation of creatinine levels among subjects who died within 5 years?

n	mean	sd
133	1.206015	0.4618408

- Units for mean is mg/dl

(c)

What are the point estimate, the estimated standard error of that point estimate, and the 95% confidence interval for the mean creatinine level in a population of similar subjects who would survive at least 5 years?

Point estimate is 1.0325 mg/dl SE is 0.009974404 95% CI is 1.01 to 1.05

(d)

What are the point estimate, the estimated standard error of that point estimate, and the 95% confidence interval for the mean creatinine level in a population of similar subjects who would die within 5 years?

Point estimate is 1.2060 mg/dl SE is 0.01885457 95% CI is 1.12 to 1.28

(e)

What are the point estimate and the 95% confidence interval for the difference in creatinine means between a population of similar subjects that survives at least 5 years and a population of similar subjects that dies within 5 years? What is the P value for testing the hypothesis that the two populations have the same mean creatinine level? What conclusions do you reach about a statistically significant association between serum creatinine and 5 year all-cause mortality?

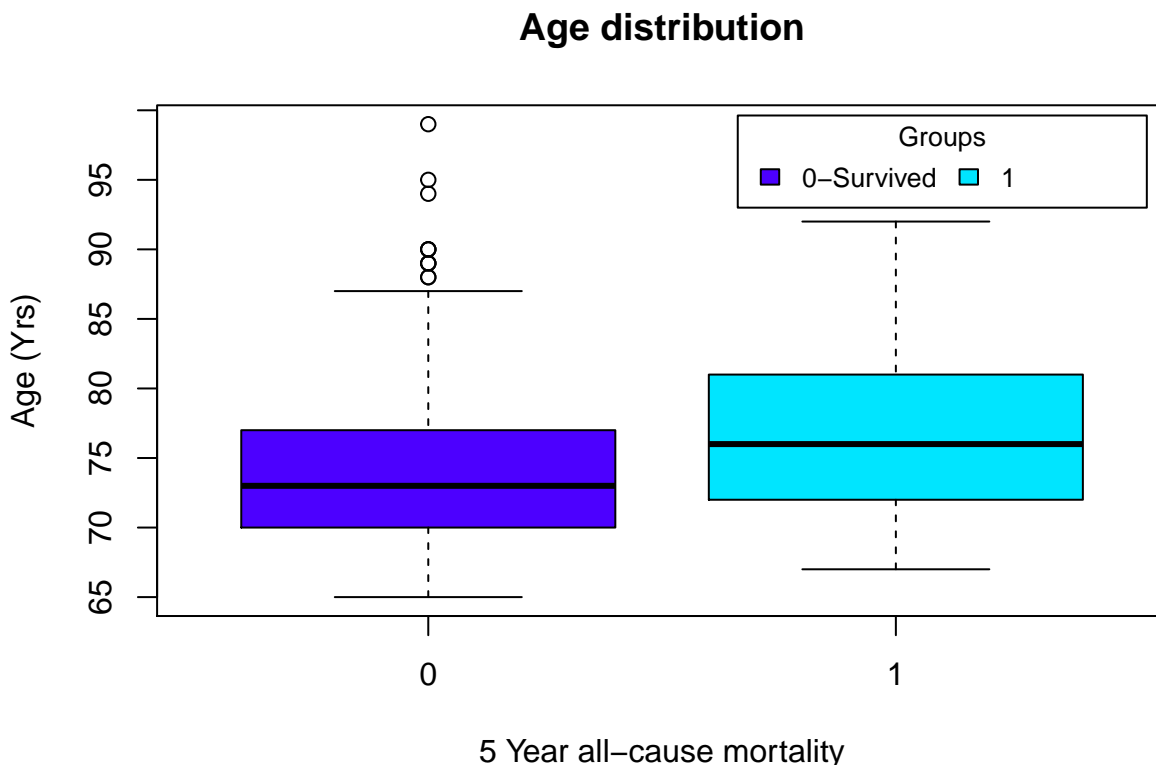
The best estimate of the difference, i.e. the point estimate is -0.173515 units (Survive Sample, Observed to Die Sample.)

95% CI , difference of mean 'crt' between the groups is (-0.25506646 , -0.09196361).

p-value is 4.501e-05, and is found to be significant, and we reject our null hypothesis that two populations have the same mean creatinine level.

(f)

Although we did not consider age at the time of enrollment in the questions above, could the association analysis for creatinine level and 5 year all-cause mortality conducted using the t-test potentially be confounded by the age of the subjects at the time of the MRI? Briefly explain why or why not this is plausible? Provide any descriptive statistics (e.g. an appropriate table, plot, etc.) giving evidence for or against the association results and conclusions above with the t-test potentially being confounded by age.



The median

Question 2

You can refer to objects you defined in previous code chunks in the body of your document's text. As an example, the standard deviation distance in the `cars` data is 25.7693775 kilometers. Note that we can use the command `round(sd(cars$dist), 1)` to round to the first decimal place, 25.8 kilometers.