Homework 5

Aji John

Winter Quarter 2018

Question 1

Perform a Poisson regression analysis to evaluate an association between 5 year all-cause mortality and creatinine by comparing the relative risk of death (or risk ratio of death) across groups defined by continuous serum creatinine level. (Only provide a formal report of inference when asked to.)

(a)

Provide an interpretation of the slope and the intercept in the Poisson regression model, and include the numerical values of the slope and intercept in your interpretation.

Exponentiated Slope is the (Rel Risk Ratio), and the exponentiated intercept is the base risk.

(b)

Give full inference for an association between 5 year all-cause mortality and serum creatinine levels from the Poisson regression model.

RR ratio between groups differing in the value of the predictor by 1 unit (i.e groups defined by the survival status) – Found by exponentiation of the slope by poisson regression: exp(b1) - 2.5641

(c)

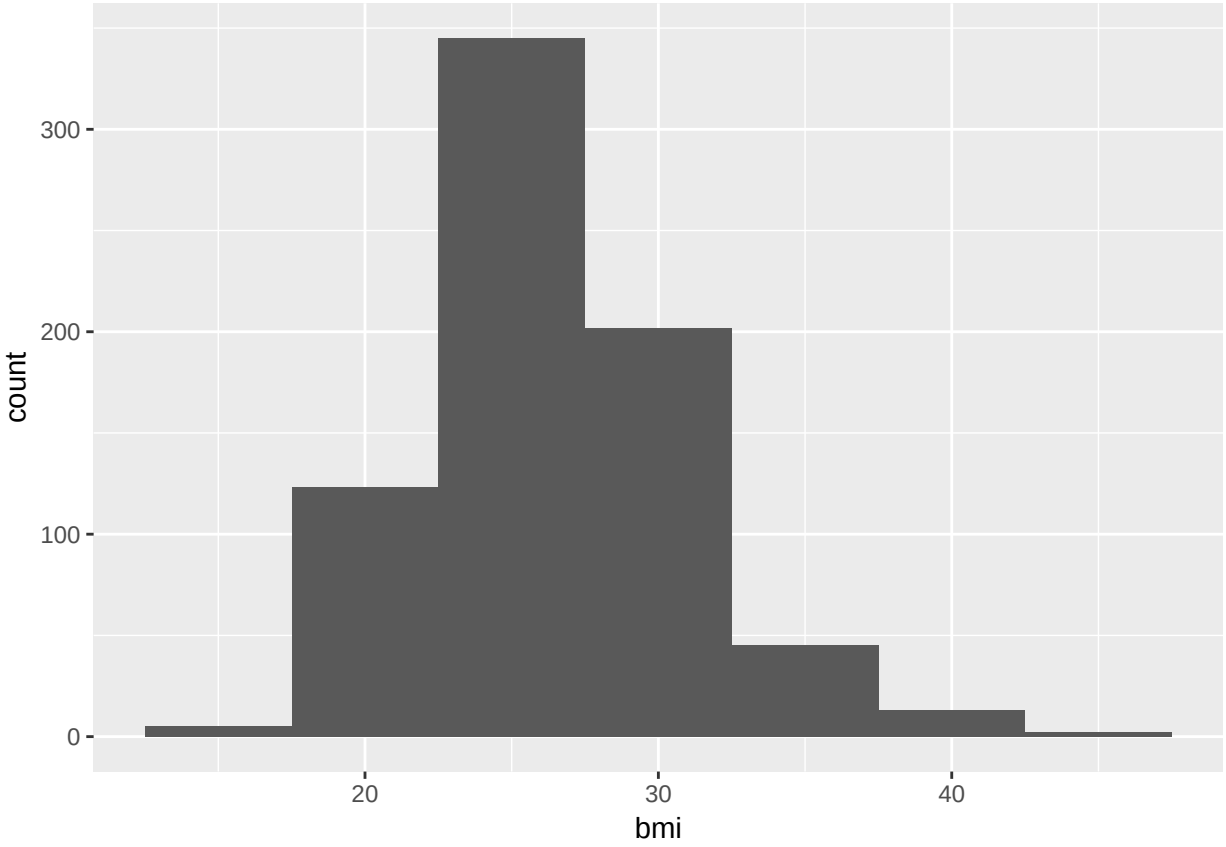
Compare the association results in part b that are based on risk ratios to using a logistic regression model where odds ratios of death within 5 years are used as the summary measure for an association with serum creatinine level (i.e., question 3 in homework 4). Briefly describe any similarities or differences in the association results.

Question 2

Questions 3 and 4 below investigate associations between serum cholesterol level, age, sex, and body mass index (BMI). In this question we will obtain some summary statistics for these variables.

(a)

Create a variable for BMI using the height and weight measurements on the subjects. [Hint: Make sure that appropriate conversions of the weight and height measures are used in the calculation of BMI]. Provide a figure illustrating the distribution of BMI in the sample.



(b)

Provide suitable descriptive statistics for serum creatinine levels, age, sex, and BMI.

todo # Male female and Total

Variable	Male	Female	All subjects
Sample size	n = 121	n = 614	n = 735
CRTs[1]	1.2 (0.3); 0.7-4	0.9 (0.3); 0.5-3.2	1.1 (0.3); 0.5-4
Age (years)[1]	74.7 (5.6); 66-99	74.4 (5.3); 65-91	74.6 (5.5); 65-99
BMI (kg/mm)[1]	26.3 (3.8); 16-42	26.5 (4.9); 15-47	26.4 (4.3); 15-47

[1] mean (sd); min-max are reported

Question 3

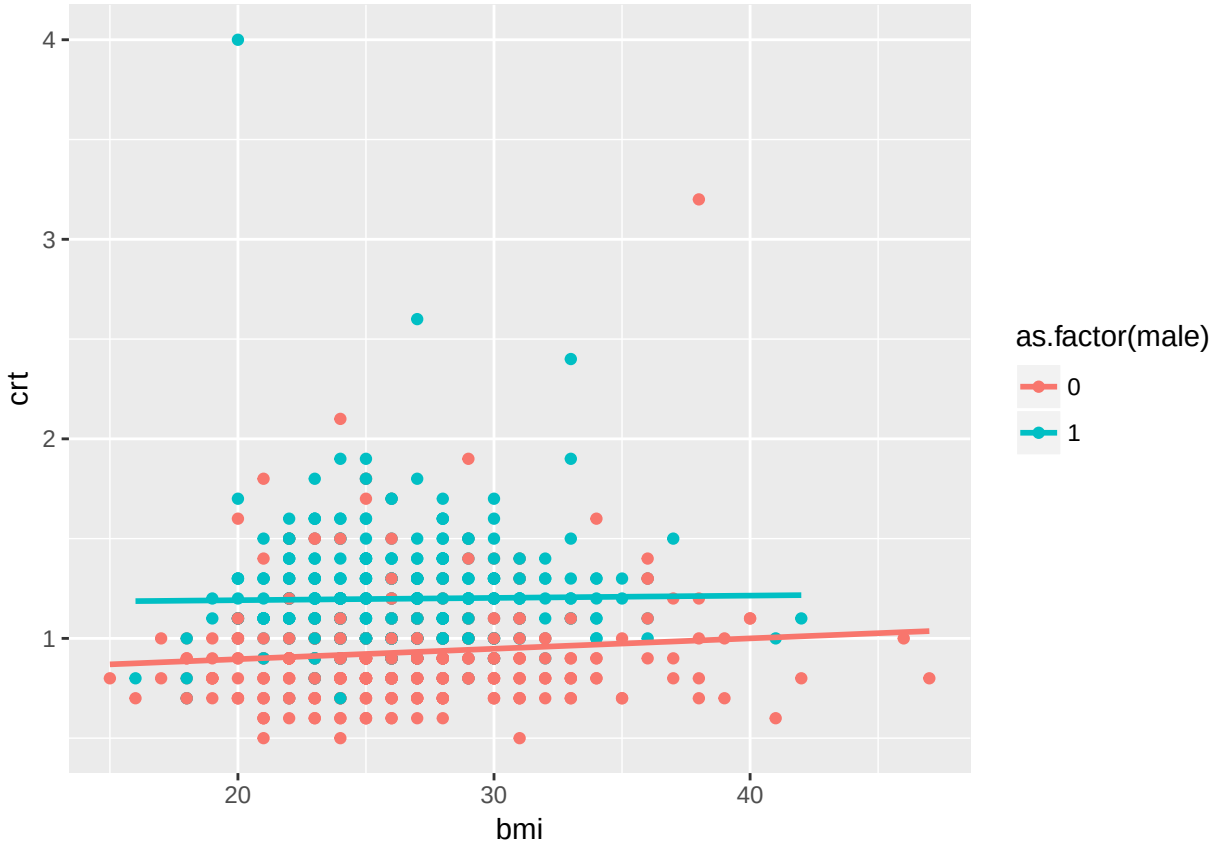
We are interested in examining how mean serum creatinine levels vary by BMI and sex. In the questions below, you do not need to provide full statistical inference. Instead, just answer the following questions.

(a)

Create a scatterplot of serum creatinine levels versus BMI. Use different symbols and/or colors for each sex group, and include LOWESS (or LOESS) curves for each sex group.

```
## Warning: Removed 2 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 2 rows containing missing values (geom_point).
```



(b)

What observations do you make from the scatterplot in part a regarding the association between serum creatinine levels and BMI?

(c)

Is there evidence from descriptive statistics (question 2) and the scatterplot in part a that sex modifies the association between serum creatinine level and BMI? Explain your reasoning.

The lines diverge, so it is an effect modifier

(d)

Is there evidence from descriptive statistics (question 1) and the scatterplot in part a that sex confounds the association between serum creatinine level and BMI? Explain your reasoning.

No, lines are not parallel for each bmi group

(e)

Perform an analysis to determine whether mean serum creatinine levels differ across sex groups. Briefly describe the analysis that you performed and clearly state the basis of your conclusion regarding an association.

Using t.test with unequal variance setting we find that it is not equal i.e. they differ across the sex groups. The p-value we find is < 0.05, and is found to be significant.

Alternatively, doing linear regression

(f)

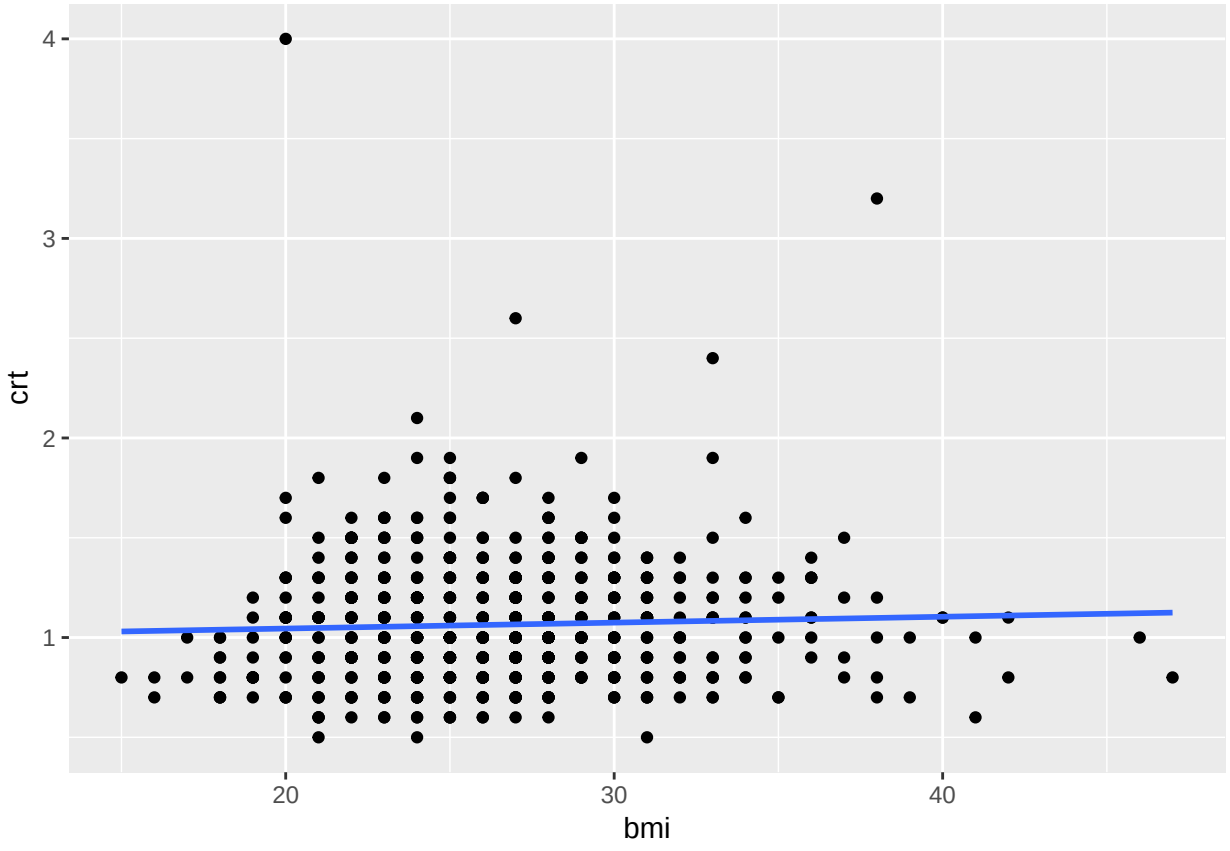
Perform an analysis to determine whether there is a linear trend in mean serum creatinine levels by BMI. Briefly describe the analysis that you performed and clearly state the basis of your conclusion regarding an association.

Question 2:

note: round() Does not print trailing zeros

```
## Warning: Removed 2 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 2 rows containing missing values (geom_point).
```



We find that slope is 3.18 * 10^-3 which at p-value of .3113 is not significant.

(g)

Perform an analysis to determine whether mean serum creatinine levels differ across sex groups after adjustment for BMI. Briefly describe the analysis that you performed and clearly state the basis of your conclusion regarding an association.

We find that after adjusting for sex, the slope is found to be still not significant (p-value 0.1928)

(h)

Perform an analysis to determine whether there is a linear trend in mean serum creatinine levels by BMI after adjustment for sex. Briefly describe the analysis that you performed and clearly state the basis of your conclusion regarding an association.

Linear trend is there as slope is .003813, but is found to be not significant(p-value 0.192).

(i)

Perform an analysis to determine if sex modifies the association between mean serum creatinine levels and BMI. Briefly describe the analysis that you performed and clearly state the basis of your conclusion regarding an association.

We used the below model (crt ~ bmi*male) to explore the effect modification. The interaction (bmi:male) is not found to be present as p-value is 0.5261.

(j)

How would you summarize the association between serum creatinine levels and BMI and sex? Provide a summary of your findings that is suitable for inclusion in a manuscript.

Question 4

Now consider a multivariate linear regression analysis with serum creatinine level as the response and the variables age, sex, and BMI as predictors.

(a)

Provide an interpretation of the intercept in the regression model. Is the slope estimate scientifically useful?

The intercept is the estimate of the mean CRT level for new born female who has 0 bmi. It is not scientifically relevant, and is not possible to have it in real life.

(b)

Give full inference for the age slope in the regression model.

The age slope is the estimated change in mean CRT for one unit change (1 yr) keeping sex and bmi constant.

(c)

Give full inference for the sex slope in the regression model.

The male slope is the estimated change in mean CRT between two sex groups by keeping age and bmi constant

(d)

Give full inference for the BMI slope in the regression model.

The bmi slope is the estimated change in mean CRT level for one unit change in bmi(mg/dl) by keeping male and age constant.