



Biostatistics: The New Frontier

Grade A. Author

Department of Quantitative Health Sciences

January 1, 2021

Just some text

Lots of people have never heard of biostatistics. Time to spread the word.

Enumerated lists

1. First things first
 - 1.1 I have a lot to say about biostatistics
 - 1.2 Just you wait!
2. And then, a second thing

Bulleted lists

- The goal is to learn about biostatistics
 - ▶ Models
 - ▶ Predictions
- Then we'll move on to other topics

Inserting a figure

Here's a picture of the Taussig Cancer Center



Words, bullets, and equations

Given some data y, x_1, x_2, \dots, x_p , we are interesting finding a likely value for y given the value of predictors $x \equiv x_1, x_2, \dots, x_p$.

- Here, y is continuous. (Called outcome, response, “dependent variable”).
- The x 's can be continuous, binary, categorical. (Called predictor, covariate, “independent variable”).
- We want $E(y|x) = f(x)$; we observe $y = f(x) + \epsilon$.

Side by side figures

On the left is incorrect mask use! Cover your nose, like on the right.



Inline equations and nested lists

A linear regression model is a particular type of parametric regression.

- Assume $f(x; \beta) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots$
- Focus is on β_0, β_1, \dots
- “Linear” refers to the β ’s, not the x ’s:
 - ▶ $f(x) = \beta_0 + \beta_1 x + \beta_2 x^2$ is a linear model
 - ▶ $f(x) = \beta_0 + x^{\beta_1}$ is not
 - ▶ $f^*(x) = \beta_0^* + \beta_1 x^*$

Code blocks using verbatim

```
> lrmod <- lm(Sepal.Length ~ Species, data = iris)
> summary(lrmod)

Call:
lm(formula = Sepal.Length ~ Species, data = iris)

Residuals:
    Min      1Q  Median      3Q     Max 
-1.6880 -0.3285 -0.0060  0.3120  1.3120 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 5.0060    0.0728  68.762 < 2e-16 ***
Speciesversicolor 0.9300    0.1030   9.033 8.77e-16 ***
Speciesvirginica 1.5820    0.1030  15.366 < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5148 on 147 degrees of freedom
Multiple R-squared:  0.6187, Adjusted R-squared:  0.6135 
F-statistic: 119.3 on 2 and 147 DF,  p-value: < 2.2e-16
```

CCF plot colors

It's easy to make plots using the Cleveland Clinic brand colors in R.

Check out this website for details: http://www.emilyzabor.com/ezfun/articles/ccf_color_palette.html.

