Review

Grinnell College

December 13, 2024

Example 1

Skittles candy has 5 flavors, we want to know if they are evenly distributed. We buy 20 packs, each containing 50 pieces. We then sort them by color and get the following results:

	Red	Purple	Yellow	Green	Orange
Count	208	222	181	192	197

▶ How would we test if these were evenly distributed?

▶ What is our null hypothesis?

▶ What do we find?

Example 2

The results of a clinical trial exploring the side effects of a drug alongside a placebo are given below

	Side Effects	None
Drug	57	143
Placebo	22	76

Does treatment appear to be associated with the onset of side effects?

Example 3

A device that can identify recent THC consumption from saliva

- ▶ What is null hypothesis?
- ▶ What constitutes a Type I and Type II error?
- ▶ Suppose that the device has a Type I error rate of 5% and a Type II error rate of 20%. Suppose further 500 drivers are stopped, with approximately 2% of these drivers being intoxicated. Assuming that the device tests *positive* for THC, what is the probability that the driver is actually inebrieted.

Example 4

Which do you think will be more impactful to add to the model that includes only Species?
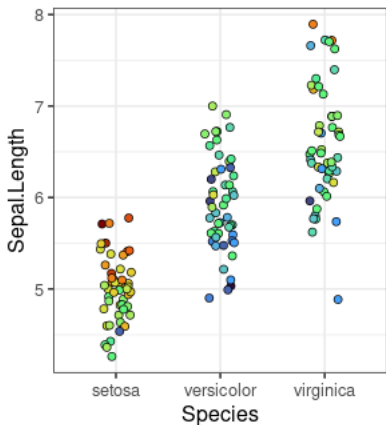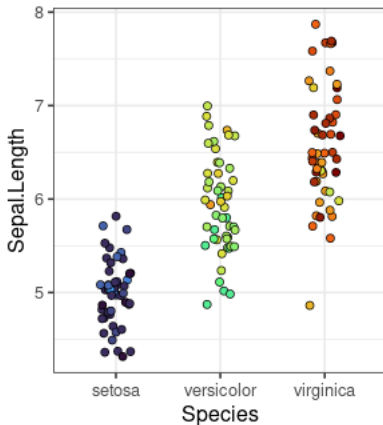
```
 1  > lm(Sepal.Length ~ Species, iris) %>% summary()
 2
 3  Coefficients:
 4                    Estimate Std. Error t value           Pr(>|t|)
 5  (Intercept)        5.0060     0.0728   68.76 < 0.0000000000000002 ***
 6  Speciesversicolor  0.9300     0.1030    9.03 0.00000000000000088 ***
 7  Speciesvirginica   1.5820     0.1030   15.37 < 0.0000000000000002 ***
 8  ---
 9  Multiple R-squared:  0.619
10
11  > lm(Sepal.Length ~ Species + Sepal.Width, iris) %>% summary()
12
13  Coefficients:
14                    Estimate Std. Error t value           Pr(>|t|)
15  (Intercept)         2.251      0.370    6.09       0.0000000095681 ***
16  Speciesversicolor   1.459      0.112   13.01 < 0.0000000000000002 ***
17  Speciesvirginica    1.947      0.100   19.47 < 0.0000000000000002 ***
18  Sepal.Width         0.804      0.106    7.56       0.0000000000042 ***
19  ---
20  Multiple R-squared:  0.726
21
22  > lm(Sepal.Length ~ Species + Petal.Width, iris) %>% summary()
23
24  Coefficients:
25                    Estimate Std. Error t value           Pr(>|t|)
26  (Intercept)        4.7804     0.0831   57.54 < 0.0000000000000002 ***
27  Speciesversicolor -0.0603     0.2304   -0.26                  0.79
28  Speciesvirginica  -0.0501     0.3582   -0.14                  0.89
29  Petal.Width        0.9169     0.1939    4.73             0.0000053 ***
30  ---
31  Multiple R-squared:  0.669
```