

Categorical data analyses

Module 11 review

Statistics 251: Statistical Methods

Checklist

- (1) State hypotheses, check assumptions if requested
- (2) State χ^2 test statistic, df , and $pvalue$ from output
- (3) State test results
- (4) Make conclusion in context from results
- (5) State possible error that could have been made and discuss it within the context

CHOCOLATE!

M&M's, made by Mars Inc., says that their population of colors of their milk chocolate M&M candies are as follows, along with a random sample of a bag of 106 candies. Is there sufficient evidence the bag of candies has the same distribution of colors as stated by Mars?

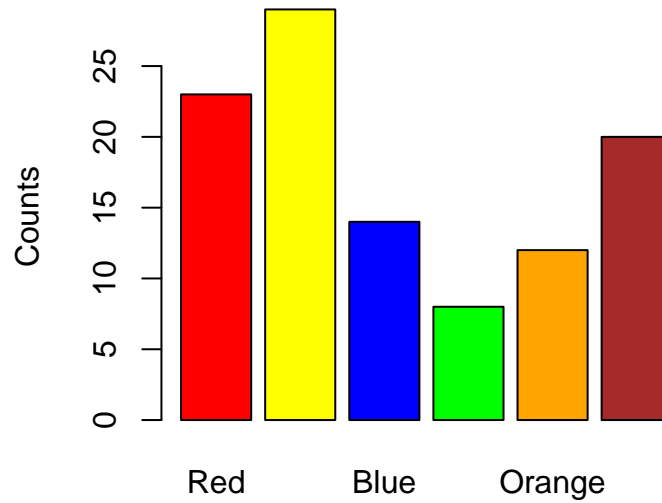
Color	Red	Yellow	Blue	Green	Orange	Brown
Population %	20%	20%	10%	10%	10%	30%
Sample count	23	29	14	8	12	20

M.M

```
colors observed mars
1 Red 23 0.2
2 Yellow 29 0.2
3 Blue 14 0.1
4 Green 8 0.1
5 Orange 12 0.1
6 Brown 20 0.3
```

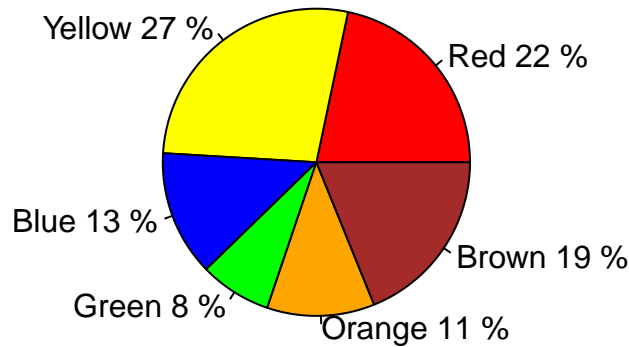
```
barplot(observed, names.arg=colors, col=gcolors, ylab='Counts')
title("Distribution of M&M Colors")
```

Distribution of M&M Colors



```
pie(observed,labels=lbls,col=gcolors,main="Distribution of M&M Colors")
```

Distribution of M&M Colors



```
chisq.test(observed,p=mars)$expected
```

```
[1] 21.2 21.2 10.6 10.6 10.6 31.8
```

```
chisq.test(observed,p=mars)
```

Chi-squared test for given probabilities

data: observed

X-squared = 9.3145, df = 5, p-value = 0.09716

H_0 : M&M colors are as stated by Mars

H_a : H_0 is not true

Assumptions:

- (1) The data must be counts from categories: yes
- (2) Independence of observations: yes
- (3) $E_i \geq 5$; individual expected values must be at least 5

Organization of information:

$df = k - 1 = 6 - 1 = 5$ where k = number of categories (candy colors)

$\alpha = 0.05$ (assumed because not specifically stated otherwise)

$\chi^2 = 9.3145$, $df = 5$, $pvalue = 0.09716$

Results: $pvalue = 0.09716 \not\leq \alpha(0.05) \therefore H_0$ is not rejected

Conclusion: since the null is not rejected, the distribution of M&M colors follow the population colors stated by Mars

Error: since H_0 was not rejected, a Type II error (failing to reject null when null is false) could have been made; we think the distribution of M&M colors follow the population colors stated by Mars but they do not.

Doh... wrong Homer

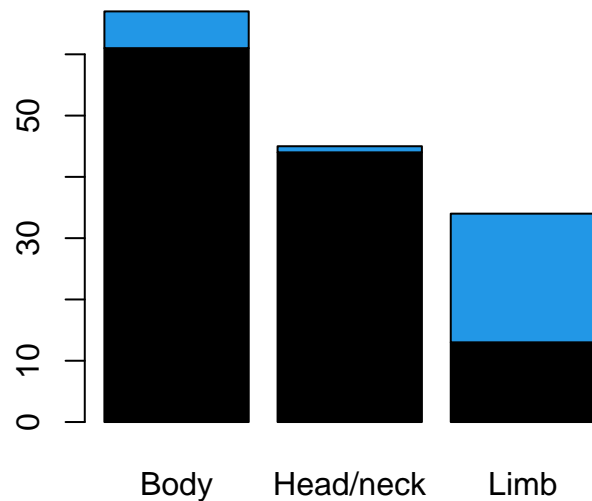
Homer's Iliad is an epic poem, compiled around 800 BCE, that describes several weeks of the last year of the 10-year siege at Troy (Ilion) by the Achaeans. The story centers on the rage of the great warrior Achilles. But it includes many details of injuries and outcomes, and is thus the oldest record of Greek medicine. Below is a table of 146 recorded injuries for which both injury site and outcome are provided in the Iliad. Is there an association between injury site and outcome of survival?

	Lethality		
Site	Fatal	Not fatal	Total
Body	61	6	67
Head/neck	44	1	45
Limb	13	21	34
Total	118	28	146

```
doh
```

```
      Lethality
Site   Fatal Not fatal
Body    61      6
Head/neck 44      1
Limb    13     21
```

```
barplot(t(doh),col=c(1,4))
```



```
chisq.test(doh)$expected
```

```
      Lethality
Site   Fatal Not fatal
Body  54.15068 12.849315
Head/neck 36.36986  8.630137
Limb    27.47945  6.520548
```

```
chisq.test(doh)
```

Pearson's Chi-squared test

```
data: doh
X-squared = 52.647, df = 2, p-value = 3.698e-12
```

H_0 : Wound site and survival were independent at the Battle of Troy

H_a : H_0 is not true

Assumptions:

- (1) The data must be counts from categories: yes
- (2) Independence of observations: yes
- (3) $E_i \geq 5$; individual expected values must be at least 5

Organization of information:

$$df = (r - 1)(c - 1) = (3 - 1)(2 - 1) = 2$$

$\alpha = 0.05$ (assumed because not specifically stated otherwise)

$$\chi^2 = 52.647, df = 2, pvalue = 3.698e-12 = 3.698 \times 10^{-12} \approx 0$$

Results: $pvalue \approx 0 \leq \alpha(0.05) \therefore H_0$ is rejected

Conclusion: since the null is rejected, there is sufficient evidence that there is an association (they are dependent) between injury site and outcome of survival

Error: since H_0 was rejected, a Type I error (reject null when null is true) could have been made; we think there is an association between injury site and outcome of survival but there is not