

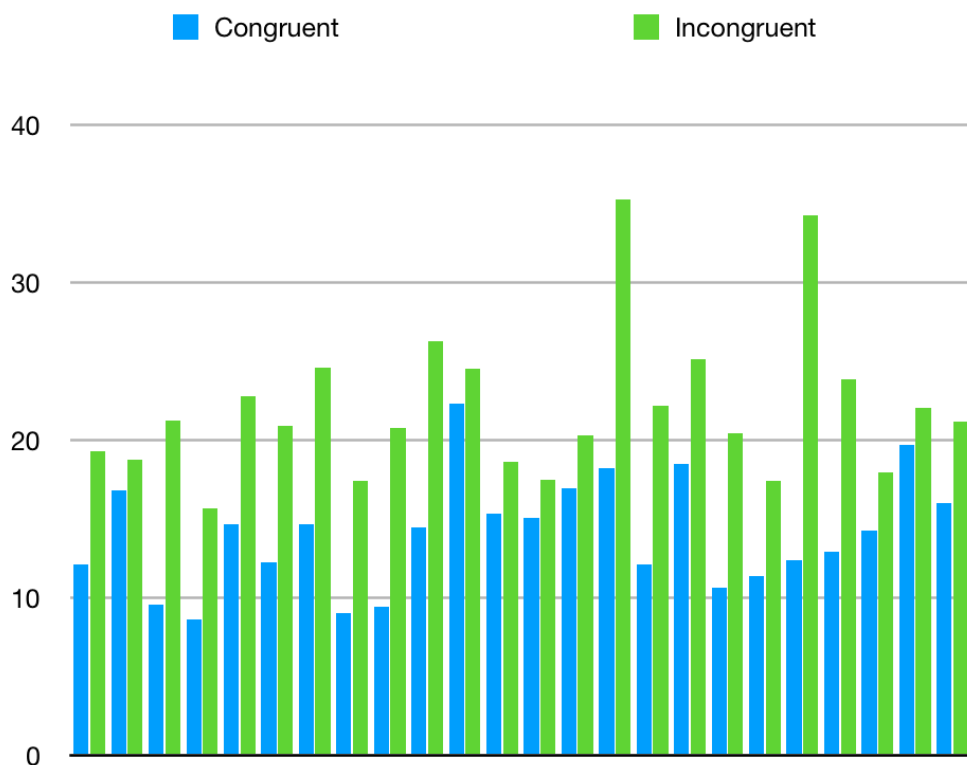
## 调查问题

---

1. 我们的**自变量**是显示的文字和其显示的颜色是否一致，**因变量**是说出同等大小的列表中的墨色名称的时间。
2. **零假设**为在一致文字条件和不一致文字条件时，说出墨色名称的时间是一样的，即假设条件 $H_0$ ，一致文字条件使用的时间总体均值 $\mu_1$ ，等于不一致文字条件使用的时间总体均值 $\mu_2$ ，记作 $H_0: \mu_1 = \mu_2$ 。**对立假设**为在一致文字条件时，相比不一致文字条件，说出墨色名称的时间是不同的，设置假设条件 $H_A$ ，一致文字条件使用的时间总体均值 $\mu_1$ ，不等于不一致文字条件使用的时间总体均值 $\mu_2$ ，记作 $H_A: \mu_1 \neq \mu_2$ 。

这里使用**相依样本t检验**。由于要求“每位参与者必须全部完成并记录每种条件下使用的时间”，可知数据之间相依关系，并不是独立的。使用t检验是由于总体参数未知，不可以通过样本估计出来，并且样本数量只有24，小于30，很难判断是不是正态分布，t检验的优点在于，只要正态假设未被严重违反，对结果的影响都不会很大。（参考地址：[Udacity论坛讨论](#)）单尾检验，因为预测具有方向性，不一致文字条件下，说出墨色名称的时间预计会变长，一致性文字时间相对会短。

3. 两组数据的样本数量分别为24，**一致的条件下**，均值为14.05，标准差为3.559；**不一致的条件下**，均值为22.02，标准差为4.797。
4. 样本数据显示的柱状图如下：



可以看出不一致的情况下，所需要的时间更多。

5.

- 样本量 $N=24$ ；
- 显著性水平 $\alpha=0.05$ ；
- 自由度 $df=23$ ；
- $t$ 临界值 $t\text{-critical}=2.069$ ；
- 样本差值的均值 $MD=14.05-22.2=-7.97$ ；
- 样本差值的标准差 $SD=4.865$ ；
- 样本差值的标准误差 $SE=SD/\sqrt{n}=4.865/\sqrt{24}=0.993$ ；
- $t$ 统计量 $t\text{-statistic}=MD/SE=-7.97/0.993=-8.026$ ；
- 检验所得的 $P$ 值，根据 $t$ 统计量为 $-8.026$ 和自由度 $23$ ，可知 $p<\alpha$ 。

计算总体均值差异的置信区间 $CI$ ： $(MD - t\text{-critical} * SE, MD + t\text{-critical} * SE)$ ，结果为 $(-10.025, -5.915)$ 。 $t$ 临界值是 $2.069$ ，总体均值差异没有落在95%的置信区间内，所以拒绝接受零假设，得出在一致文字条件下，相比不一致文字条件时，说出墨色名称的时间更短。