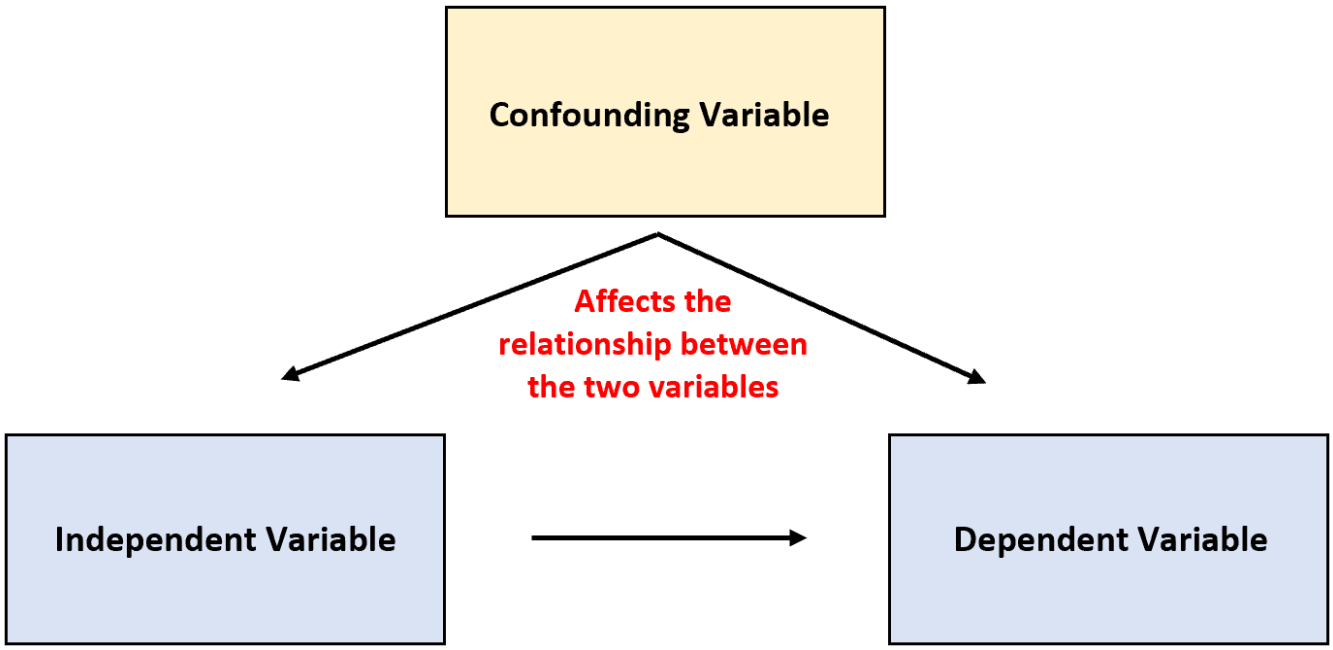
91. What is a confounding variable?

Ans.

A confounding variable in statistics is an 'extra' or 'third' variable that is associated with both the dependent variable and the independent variable, and it can give a wrong estimate that provides useless results.

For example, if we are studying the effect of weight gain, then lack of workout will be the independent variable, and weight gain will be the dependent variable. In this case, the amount of food consumption can be the confounding variable as it will mask or distort the effect of other variables in the study. The effect of weather can be another confounding variable that may later the experiment design.



92. What are the steps we should take in hypothesis testing?

Ans.

1. State the null hypothesis
2. State the alternate hypothesis
3. Which test and test statistic to be performed
4. Collect Data
5. Calculate the test statistic
6. Construct Acceptance / Rejection regions
7. Based on steps 5 and 6, draw a conclusion about H0

93. How would you describe what a 'p-value' is to a non-technical person or in a layman term?

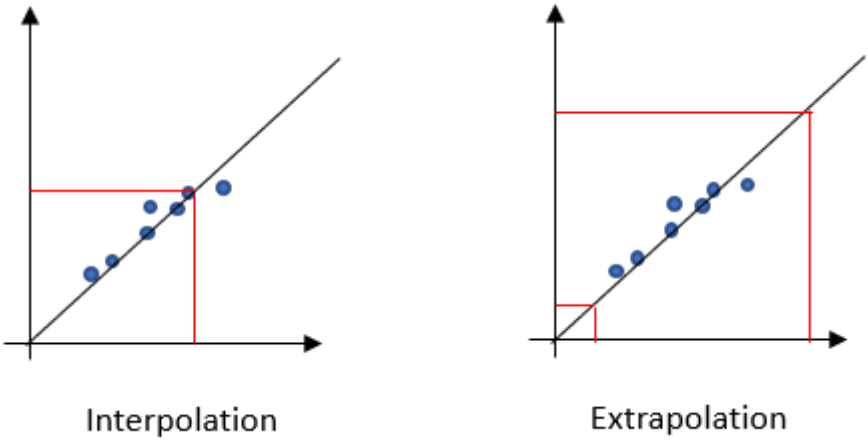
Ans.

The best way to describe the p-value in simple terms is with an example. In practice, if the p-value is less than the alpha, say of 0.05, then we're saying that there's a probability of less than 5% that the result could have happened by chance. Similarly, a p-value of 0.05 is the same as saying "5% of the time, we would see this by chance."

94. What does interpolation and extrapolation mean? Which is generally more accurate?

Ans.

- Interpolation is a prediction made using inputs that lie within the set of observed values.
- Extrapolation is when a prediction is made using an input that's outside the set of observed values.
- Generally, interpolations are more accurate.



95. What is an inlier?

Ans.

An inlier is a data observation that lies within the rest of the dataset and is unusual of an error. Since it lies in the dataset, it is typically harder to identify than an outlier.

96. You flipped a biased coin (p(head)=0.8) five times. What's the probability of getting three or more heads?

Ans.

To start off the question, we need 3, 4, or 5 heads to satisfy the cases.

5 heads: All heads, so (4/5)^5=1024/3125

4 heads: All heads but 1. There are 5 ways to organize this, and then a (4/5)^4*(1/5)^1=256/3125. Since there are 5 cases, we have 1280/3125.

3 heads: All heads but 2. There are 10 ways to organize this, and then a (4/5)^3*(1/5)^2=64/3125. Since there are 10 cases, we have 640/3125.

We sum all these cases up to get (1024+1280+640)/3125=2944/3125.

We have a 2944/3125 or 0.94208 probability to get 3 or more heads.

97. Infection rates at a hospital above a 1 infection per 100 person-days at risk are considered high. A hospital had 10 infections over the last 1787 person-days at risk. Give the p-value of the correct one-sided test of whether the hospital is below the standard.

Ans.

The test the hypothesis H0:λ=0.01

Ha:λ<0.01.

We have X=10, t=1787 and we assume that XH0∼Poisson(λ.t).

rate <- 1/100

pval <- ppois(10, lambda = rate * t)

round(pval, 2) #0.03

98. In a population of interest, a sample of 9 men yielded a sample average brain volume of 1,100cc and a standard deviation of 30cc. What is a 95% Student's T confidence interval for the mean brain volume in this new population?

Ans.

mu = 1100

sigma = 30

quantile = 0.975 # 95% with 2.5% both side

CI = mu + c(-1, 1) * sigma * qt(quantile, df=n-1)/sqrt(n)

99. What Chi-square test?

Ans.

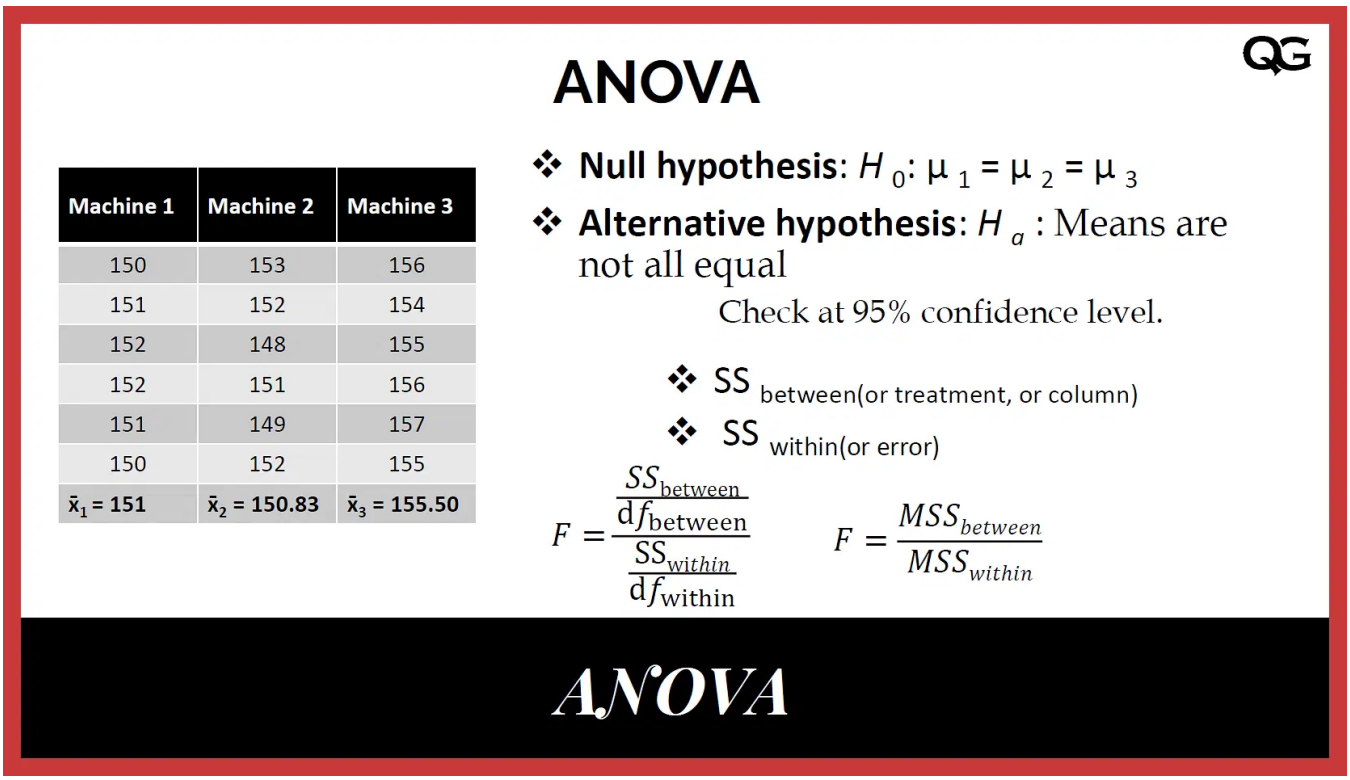
A statistical method is used to find the difference or correlation between the observed and expected categorical variables in the dataset. Example: A food delivery company wants to find the relationship between gender, location and food choices of people in India. It is used to determine whether the difference between 2 categorical variables is:

● Due to chance or ● Due to relationship

100. What is the ANOVA test?

Ans.

An ANNOVA (analysis of variance) is a statistical test used to determine statistical difference between two or more categorical groups by calculating mean and variance.



In [ ]:

```
1
```

Thanks

Github: https://github.com/saisubhasish (https://github.com/saisubhasish)

In [ ]:

```
1
```