*** What does a z-value express?**

z-score is distance from mean in standard deviation units.

Compares the distribution of data.

- Is related to standard deviation.
- Calculation:
    - Each data point minus the average data point. Divided by deviation.

Expresses how many standard deviations an observation/data point is away from the mean (of the population?).

It's a standardized value - so can be used to compare to different datasets for examples.

A standardized score, specifically scaled to be on a distribution with a mean of 0 and a standard deviation of 1.

It expresses how many standard deviations a data point is away from the mean. It allows us to calculate the probability of a specific value occuring within our distribution.

It's a standardized value, meaning that we can use it to compare several data sets.

We derive it from a distribution of the data with the z-values on the x-axis centered at zero (y axis is the probability of observing a specific set value). It is a relation between 2 things; standard error (the noise in the data) & the sample mean.

The **standard error** is the sample mean divided by sqrt(n).

z value indicates number of standard deviations from the mean of 0, which is the mean of a standard normal distribution.

*** What does a p-value express?**

The p-value is the propability of obtaining a sample given the null hypothesis is true.

Used to evaluate and reject hypotheses.

    - p can never reject the actual hypothesis.
- Is derived from (infinite number of) t-tests.
- If a p-value is below .05, we reject the null hypothesis.
    - In the long run, only 5% of the time, this rejection will be false.

Alpha expresses the probability of a type I error (i.e., rejecting the null hypothesis when it is true).

Under the assumption that the null hypothesis is true, the probabliity of obtaining a result that is the same or more extreme.

The probability of getting your obtained value (or more extreme) under the null-hypothesis

The probability of observing a score as the current one or more extreme given the null hypothesis is true. We say that the p-value is significant if it is less than alpha.

p-values (usually .05) are either/or, not gradual, a lower p-value does not suggest something is more likely, and a large p-value does not suggest something is less likely. It accepts that 5% of results are a result of type1 errors (false positives)

*** What is the relation between a t-value and a z-value?**

No idea

What value one is using depends on your given/collected data. One fits better for certain sample sizes than the other.

    - z-value: population mean
    - t-value: sample mean
- In the end, both will be used for calculating a p-value.

Both show how many standard deviations a data point is away from the mean

z-value is for when you know the mean and variance of the population

t-values are for when you don't know the mean of the population or when the sample size is very small.

t and z are the same except for the degress of freedom

?

t-value, is the same, except you don’t have enough data to test assumptions, so it is more restrictive as to show significance. It has one less degree of freedom.

*** Explain symbols from the general linear model and ordinary least squares**

Linear model: Y = vector of target values, X = design matrix (containing x-values and 1's (size of matrix is dependent on number of parameters and dimensions)), beta = a vector of parameter values, epsilon = the error

Linear model: Y = vector of target values, X = design matrix (containing x-values and 1's (size of matrix is dependent on number of parameters and dimensions)), beta = a vector of parameter values, epsilon = the error

Y = output

    - X = input
    - Beta = slope of model
    - Epsilon = errors / noise
- 2nd one: the estimation of beta (ie. the slope of the model)
    - ^ = estimation.
    - T = transposition of the matrix.

The first equation is the formula for the general linear model:

Y = Outcome variable (matrix)

X = Dependent variable (matrix)

Beta = Coefficients (vector)

Episilon = Error (vector)

Beta Hat = estimated coefficient (vector)

The second equation is the OLS (Ordinary Least Squares) estimator, where the first part is the design matrix (with X-values) and the second part is the Y-values.

X^t = X transposed

X^-1 = X inverted

- Y = X*beta +e

Y = predicted values

Beta = coefficient estimates

e = error

X = a design matrix (with observed data points)

Beta-hat:

Beta-hat: estimated beta values (slope and intercept). A vector

X^t = design matrix transposed ****

X = design matrix

(^-1) = inversed

Y = obtained y values

Y: A vector of outcome values.

X: The design matrix. The number of columns is equal to the number of predictors in our model plus the constant term.

Beta: A vector of our regression estimates.

epsilon: A vector of errors.

The second equation is how you calculate the ordinary least squares estimator.

4.1 First equation

Y: The linear model. Y is the vector of our dependent variable

X: Our independent variable, design matrix

Beta: Our coefficient vector/parameter values. How much weight should we assign to our independent variables/predictors? Slope and intercept or more, if you have several predictors)

Error: Error - what is not explained by our model

4.2 Last equation - ordinary least squares estimator

Beta_hat: vector of our estimated coefficients/parameters

$\mathbf{X}^{\mathbf{T}}$: Transpose of the desig matrix

X: Design matrix

-1: the inverse