Null Hypothesis Significance Testing (NHST)

BIOS 6611

CU Anschutz

Week 4

1. **Fisher vs. Neyman-Pearson**

2. **Null Hypothesis Significance Testing (NHST)**

Fisher vs. Neyman-Pearson

Fisher Review

Fisher's approach to testing data focused on:

- permutation test
- calculation of a "p-value"
- only defining a null hypothesis
- comparisons done *a posteriori*

Neyman-Pearson Review

Neyman and Pearson's approach to testing data focused on:

- explicitly defining H_0 and H_1
- using rejection regions instead of p-values
- repeated sampling is assumed for properties like type I (α) and type II (β) errors
- α and β should be defined *a priori* and a study designed based on these assumptions with adequate power

And the Winner Is. . .

So who ultimately won the statistical fight to have their methods used?

Null Hypothesis Significance Testing (NHST)

NHST

Instead of strictly using either Fisher's approach or the Neyman-Pearson approach to data analysis, we have created an amalgamated monster from their combined ideas.

For example, we often design studies based on α (type I error) and $1 - \beta$ (power), but then evaluate our results with a combination of p-values and confidence intervals.

Specifically considering p-values:

- **Fisher**: The probability of obtaining a result as extreme or more extreme than the one observed in the sample under the null distribution. This is usually obtained from a *parametric sampling distribution* that the *test statistic* for the data is assumed to follow.
- **Neyman-Pearson**: The α-level that would have had to have been specified to just (barely) reject H_0 based on the observed data.

An NHST Example

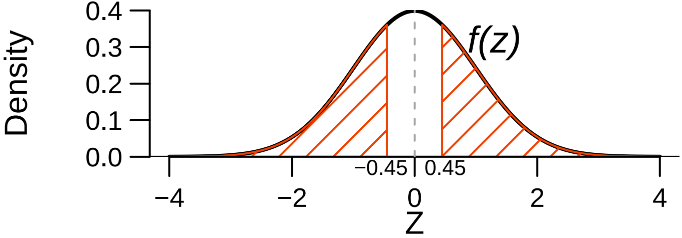
If measure cholesterol levels where $n = 12$ and $\sigma^2 = 46^2$ $(\text{mg/dL})^2$, what is the *probability* of observing an $\bar{X}$ value as or more extreme than 217 mg/dL if the true value is $H_0 : \mu = 211$ mg/dL?

$$P(|\bar{X} - 211| > |217 - 211| \Big| H_0 : \mu = 211 \text{ mg/dL})$$

$$P\left(\left|\frac{\bar{X} - 211}{46/\sqrt{12}}\right| > \frac{6}{46/\sqrt{12}} \Bigg| H_0 : \mu = 211 \text{ mg/dL}\right)$$

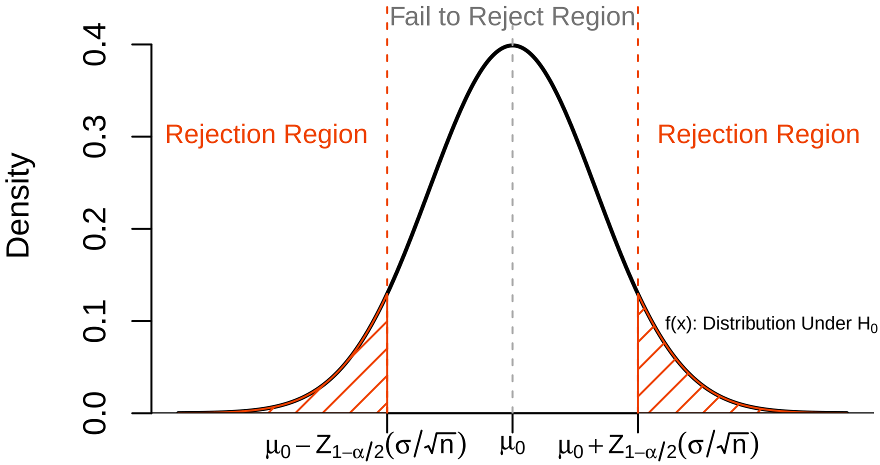
$$P(|Z| > 0.45) = (1 - 0.6736) \times 2 = 0.6528$$

Conclusion: p>0.05, so there is not enough evidence to reject H_0.



Confidence Intervals and Hypothesis Tests

For the hypothesis test $H_0 : \mu = \mu_0$ vs. $H_1 : \mu \neq \mu_0$, we define the rejection regions as



Confidence Intervals and p-values

Given the amalgamation of p-values and α, we can draw a direct connection between the p-value and the confidence interval. If $p < \alpha$ (i.e., we reject H_0), then the $(1 - \alpha)$% confidence interval will *exclude* the null value.

A benefit of confidence intervals over p-values alone is that we can understand the variability of our test statistic.

NHST Conclusion and Alternatives

NHST is the predominant approach to conducting statistical tests, but it is not without controversy.

There are also alternative approaches or paradigms (e.g., Bayesian).

We will be primarily working in the NHST framework the rest of the semester for 6611.