Government 10: Quantitative Political Analysis

Sean Westwood

Statistical Significance

How to make sense of results?

So far:

How to make sense of results?

So far:

- Focus on findings and computing differences.

How to make sense of results?

So far:

- Focus on findings and computing differences.
- but do these differences matter?

How to make sense of results?

So far:

- Focus on findings and computing differences.
- but do these differences matter?
  - How would we know?

How to make sense of results?

So far:

- Focus on findings and computing differences.
- but do these differences matter?
  - How would we know?

Today:

How to make sense of results?

So far:

- Focus on findings and computing differences.
- but do these differences matter?
  - How would we know?

Today:

- Systematic approach for understanding which differences are meaningful and which are not.

What is Statistical Significance?

Statistical significance is a measure that helps us determine whether the observed results in a study are likely due to chance or if they reflect a true effect.

What is Statistical Significance?

Statistical significance is a measure that helps us determine whether the observed results in a study are likely due to chance or if they reflect a true effect.

It helps researchers decide if their findings are reliable and can be generalized to a larger population.

Important Considerations

1. Statistical significance does not imply practical significance

Important Considerations

1. Statistical significance does not imply practical significance
2. Large differences are not always (or even often) significant

Important Considerations

1. Statistical significance does not imply practical significance
2. Large differences are not always (or even often) significant
3. Larger samples provide more reliable results

Important Considerations

1. Statistical significance does not imply practical significance
2. Large differences are not always (or even often) significant
3. Larger samples provide more reliable results
4. Larger samples can make EVERYTHING significant

Important Considerations

1. Statistical significance does not imply practical significance
2. Large differences are not always (or even often) significant
3. Larger samples provide more reliable results
4. Larger samples can make EVERYTHING significant
5. Significance does not imply causation

Hypothesis Testing

In the Chechnya example, we can reframe our study as a *hypothesis* test.

Hypothesis Testing

In the Chechnya example, we can reframe our study as a *hypothesis* test.

Possible Hypothesis:

Hypothesis Testing

In the Chechnya example, we can reframe our study as a *hypothesis* test.

Possible Hypothesis:
- "Artillery attacks decrease rebel activity by 5"

Hypothesis Testing

In the Chechnya example, we can reframe our study as a *hypothesis* test.

Possible Hypothesis:
- "Artillery attacks decrease rebel activity by 5"

But we need a formal framework to evaluate this hypothesis:

Hypothesis Testing

In the Chechnya example, we can reframe our study as a *hypothesis* test.

Possible Hypothesis:
- "Artillery attacks decrease rebel activity by 5"

But we need a formal framework to evaluate this hypothesis:

We call our starting hypothesis *the null hypothesis* (H_0)

Hypothesis Testing

In the Chechnya example, we can reframe our study as a *hypothesis* test.

Possible Hypothesis:
- "Artillery attacks decrease rebel activity by 5"

But we need a formal framework to evaluate this hypothesis:

We call our starting hypothesis *the null hypothesis* (H_0)

- usually framed in the negative

Hypothesis Testing

In the Chechnya example, we can reframe our study as a *hypothesis* test.

Possible Hypothesis:
- "Artillery attacks decrease rebel activity by 5"

But we need a formal framework to evaluate this hypothesis:

We call our starting hypothesis *the null hypothesis* (H_0)

- usually framed in the negative
- determine whether we can reject it based on our data.

Hypothesis Testing

In the Chechnya example, we can reframe our study as a *hypothesis* test.

Possible Hypothesis:
- "Artillery attacks decrease rebel activity by 5"

But we need a formal framework to evaluate this hypothesis:

We call our starting hypothesis *the null hypothesis* (H_0)

- usually framed in the negative
- determine whether we can reject it based on our data.

Here it would be:

-"Artillery attacks have no effect on rebel activity"

Hypothesis Testing

Notation:

Hypothesis Testing

Notation:

What we expect:

$$H_a(\text{Alternative Hypothesis})$$

Hypothesis Testing

Notation:

What we expect:

$$H_a(\text{Alternative Hypothesis})$$

What we test:

$$H_0(\text{Null Hypothesis})$$

Hypothesis Testing

Directional and broad hypotheses are possible

Hypothesis Testing

Directional and broad hypotheses are possible

Examples:

Hypothesis Testing

Directional and broad hypotheses are possible

Examples:

- “Artillery attacks decreased rebel activity”

Hypothesis Testing

Directional and broad hypotheses are possible

Examples:

- "Artillery attacks decreased rebel activity"
- "Artillery attacks increase rebel activity"

Hypothesis Testing

Directional and broad hypotheses are possible

Examples:

- “Artillery attacks decreased rebel activity”
- “Artillery attacks increase rebel activity”
- “Artillery attacks changed rebel activity”

Hypothesis Testing

Directional and broad hypotheses are possible

Examples:

- "Artillery attacks decreased rebel activity"
- "Artillery attacks increase rebel activity"
- "Artillery attacks changed rebel activity"

Corresponding nulls:

Hypothesis Testing

Directional and broad hypotheses are possible

Examples:

- “Artillery attacks decreased rebel activity”
- “Artillery attacks increase rebel activity”
- “Artillery attacks changed rebel activity”

Corresponding nulls:

- “Artillery attacks increased/had no effect on rebel activity”

Hypothesis Testing

Directional and broad hypotheses are possible

Examples:

- "Artillery attacks decreased rebel activity"
- "Artillery attacks increase rebel activity"
- "Artillery attacks changed rebel activity"

Corresponding nulls:

- "Artillery attacks increased/had no effect on rebel activity"
- "Artillery attacks decreased/had no effect on rebel activity"

Hypothesis Testing

Directional and broad hypotheses are possible

Examples:

- "Artillery attacks decreased rebel activity"
- "Artillery attacks increase rebel activity"
- "Artillery attacks changed rebel activity"

Corresponding nulls:

- "Artillery attacks increased/had no effect on rebel activity"
- "Artillery attacks decreased/had no effect on rebel activity"
- "Artillery attacks had no effect on rebel activity"

What do we do with a null hypothesis?

Our data will allow us to say if we have evidence to support our hypotheses or not.

What do we do with a null hypothesis?

Our data will allow us to say if we have evidence to support our hypotheses or not.

If we find significant evidence for a relationship and we are predicting a relationship, then we can reject the null hypothesis of no relationship.

What do we do with a null hypothesis?

We can never say:

What do we do with a null hypothesis?

We can never say:

1. Our hypothesis is true/correct.

What do we do with a null hypothesis?

We can never say:

1. Our hypothesis is true/correct.
2. The null hypothesis is false.

What do we do with a null hypothesis?

We can never say:

1. Our hypothesis is true/correct.
2. The null hypothesis is false.

Instead, we can say:

What do we do with a null hypothesis?

We can never say:

1. Our hypothesis is true/correct.
2. The null hypothesis is false.

Instead, we can say:

1. There is (is not) significant evidence to support our hypothesis.

What do we do with a null hypothesis?

We can never say:

1. Our hypothesis is true/correct.
2. The null hypothesis is false.

Instead, we can say:

1. There is (is not) significant evidence to support our hypothesis.
2. We can reject (fail to reject) the null hypothesis

What do we do with a null hypothesis?

So, we either

What do we do with a null hypothesis?

So, we either

- Reject the null and find evidence supporting the alternative hypothesis

What do we do with a null hypothesis?

So, we either

- Reject the null and find evidence supporting the alternative hypothesis

OR

What do we do with a null hypothesis?

So, we either

- Reject the null and find evidence supporting the alternative hypothesis

OR

- Fail to reject the null and do not find evidence supporting the alternative hypothesis

The next step: standard of evidence

How do we know what to make of our hypothesis and the null hypothesis?

The next step: standard of evidence

How do we know what to make of our hypothesis and the null hypothesis?

We have already looked at p-values in regression, but what do they mean?

The next step: standard of evidence

How do we know what to make of our hypothesis and the null hypothesis?

We have already looked at p-values in regression, but what do they mean?

A p-value (or probability value) is a statistical metric used to evaluate the strength of evidence against a null hypothesis in hypothesis testing.

The next step: standard of evidence

How do we know what to make of our hypothesis and the null hypothesis?

We have already looked at p-values in regression, but what do they mean?

A p-value (or probability value) is a statistical metric used to evaluate the strength of evidence against a null hypothesis in hypothesis testing.

Specifically, the p-value represents the probability of obtaining test results at least as extreme as the observed results, assuming that the null hypothesis is true.

What do we do do we know what to make of a p-value?

How do we know if p-value indicates significance or not?

What do we do do we know what to make of a p-value?

How do we know if p-value indicates significance or not?

That is, how do we know if a result is worth taking seriously?

What do we do do we know what to make of a p-value?

How do we know if p-value indicates significance or not?

That is, how do we know if a result is worth taking seriously?

We impose a threshold.

What do we do do we know what to make of a p-value?

How do we know if p-value indicates significance or not?

That is, how do we know if a result is worth taking seriously?

We impose a threshold.

It is arbitrary, but consistent

What do we do do we know what to make of a p-value?

How do we know if p-value indicates significance or not?

That is, how do we know if a result is worth taking seriously?

We impose a threshold.

It is arbitrary, but consistent

If the result of the statistical test meets the established threshold:

What do we do do we know what to make of a p-value?

How do we know if p-value indicates significance or not?

That is, how do we know if a result is worth taking seriously?

We impose a threshold.

It is arbitrary, but consistent

If the result of the statistical test meets the established threshold:

- then we have significance

What do we do do we know what to make of a p-value?

How do we know if p-value indicates significance or not?

That is, how do we know if a result is worth taking seriously?

We impose a threshold.

It is arbitrary, but consistent

If the result of the statistical test meets the established threshold:

- then we have significance

If the result of the statistical test does not meet the established threshold:

What do we do do we know what to make of a p-value?

How do we know if p-value indicates significance or not?

That is, how do we know if a result is worth taking seriously?

We impose a threshold.

It is arbitrary, but consistent

If the result of the statistical test meets the established threshold:

- then we have significance

If the result of the statistical test does not meet the established threshold:

- then we do not have significance

Where does this threshold come from?

Sir Ronald A. Fisher

Fisher recognized that real-world data comes with natural variability

Where does this threshold come from?

Sir Ronald A. Fisher

Fisher recognized that real-world data comes with natural variability

Differences in outcomes could be due to either random variation or a true effect of one variable on another.

Where does this threshold come from?

Sir Ronald A. Fisher

Fisher recognized that real-world data comes with natural variability

Differences in outcomes could be due to either random variation or a true effect of one variable on another.

To differentiate between the two, he needed a way to quantify the likelihood of observed data under a given assumption.

Where does this threshold come from?

Sir Ronald A. Fisher

Fisher recognized that real-world data comes with natural variability

Differences in outcomes could be due to either random variation or a true effect of one variable on another.

To differentiate between the two, he needed a way to quantify the likelihood of observed data under a given assumption.

Created the p-value

Where does this threshold come from?

- Originally a "rough guide" for the strength of evidence

Where does this threshold come from?

- Originally a “rough guide” for the strength of evidence
- Jerzy Neyman and Egon Pearson introduced fixed significance levels (0.05, 0.01, etc.)

Where does this threshold come from?

- Originally a "rough guide" for the strength of evidence
- Jerzy Neyman and Egon Pearson introduced fixed significance levels (0.05, 0.01, etc.)
- Honestly, very arbitrary.

Where does this threshold come from?

- Originally a “rough guide” for the strength of evidence
- Jerzy Neyman and Egon Pearson introduced fixed significance levels (0.05, 0.01, etc.)
- Honestly, very arbitrary.
  - We use a .05 threshold

Where does this threshold come from?

- Originally a “rough guide” for the strength of evidence
- Jerzy Neyman and Egon Pearson introduced fixed significance levels (0.05, 0.01, etc.)
- Honestly, very arbitrary.
  - We use a .05 threshold
  - Physicists use a .0005 threshold

Where does this threshold come from?

- Originally a “rough guide” for the strength of evidence
- Jerzy Neyman and Egon Pearson introduced fixed significance levels (0.05, 0.01, etc.)
- Honestly, very arbitrary.
  - We use a .05 threshold
  - Physicists use a .0005 threshold

We use a higher threshold because of sample size limitations and cost concerns.

How do we know if something is significant?

- Three “signs” of significance

How do we know if something is significant?

- Three “signs” of significance
- All are equivalent at the 5% threshold

How do we know if something is significant?

- Three “signs” of significance
- All are equivalent at the 5% threshold

1. T-statistic $>$ 1.96 or t-statistic $<$ -1.96

How do we know if something is significant?

- Three “signs” of significance
- All are equivalent at the 5% threshold

1. T-statistic $>$ 1.96 or t-statistic $<$ -1.96
2. P-value $<$.05

How do we know if something is significant?

- Three "signs" of significance
- All are equivalent at the 5% threshold

1. T-statistic $>$ 1.96 or t-statistic $<$ -1.96
2. P-value $<$.05
3. 95% confidence intervals do *not* include 0

How do we know if something is significant?

- Three "signs" of significance
- All are equivalent at the 5% threshold

1. T-statistic $>$ 1.96 or t-statistic $<$ -1.96
2. P-value $<$.05
3. 95% confidence intervals do *not* include 0

These ideas are all connected

- A p-value is computed from a t-statistic

These ideas are all connected

- A p-value is computed from a t-statistic
- 95% confidence intervals are computed from (indirectly) t-statistics

These ideas are all connected

- A p-value is computed from a t-statistic
- 95% confidence intervals are computed from (indirectly) t-statistics

What do they mean?

The t-statistic

A statistic to evaluate a mean or a difference in means when hypothesis testing

The t-statistic

A statistic to evaluate a mean or a difference in means when hypothesis testing

- Provides information on if what we observed matches what we expected

The t-statistic

A statistic to evaluate a mean or a difference in means when hypothesis testing

- Provides information on if what we observed matches what we expected

Significant: If the t-statistic is large (positive or negative), it means the difference between the groups is unlikely to have happened by random chance.

The t-statistic

A statistic to evaluate a mean or a difference in means when hypothesis testing

- Provides information on if what we observed matches what we expected

Significant: If the t-statistic is large (positive or negative), it means the difference between the groups is unlikely to have happened by random chance.

Not significant: If the t-statistic is small, it means the difference could easily be due to random variation.

P-values

Significant: A low p-value ($<=$ .05) indicates we should reject the null hypothesis—there might be a real effect or difference.

P-values

Significant: A low p-value ($<=$ .05) indicates we should reject the null hypothesis—there might be a real effect or difference.

Not significant: A high p-value ($>$.05) suggests that there's not enough evidence to reject the null hypothesis, and any observed difference might be due to random chance.

Confidence intervals are more complex

Assume we can not measure attitudes or features of everyone and that we will rely on a sample.

Confidence intervals are more complex

Assume we can not measure attitudes or features of everyone and that we will rely on a sample.

- With a sample we can be wrong!

Confidence intervals are more complex

Assume we can not measure attitudes or features of everyone and that we will rely on a sample.

- With a sample we can be wrong!
- There will be error.

Confidence intervals are more complex

Assume we can not measure attitudes or features of everyone and that we will rely on a sample.

- With a sample we can be wrong!
- There will be error.
- But how close did we get?

Confidence intervals are more complex

Assume we can not measure attitudes or features of everyone and that we will rely on a sample.

- With a sample we can be wrong!
- There will be error.
- But how close did we get?

A 95% confidence interval gives a range of values that, if we were to repeat the sampling process many times, would contain the true value of the parameter we are estimating in 95% of those intervals.

How to think about a confidence interval

- Assume we can repeat a survey 100 times.

How to think about a confidence interval

- Assume we can repeat a survey 100 times.
- Assume we compute a confidence interval for each survey

How to think about a confidence interval

- Assume we can repeat a survey 100 times.
- Assume we compute a confidence interval for each survey
- 95% of those intervals would contain the true population mean.

How do we compute a confidence interval

Two components:

How do we compute a confidence interval

Two components:

1. A mean, coefficient, or mean difference

How do we compute a confidence interval

Two components:

1. A mean, coefficient, or mean difference
2. A measure of error

How do we compute a confidence interval

Two components:

1. A mean, coefficient, or mean difference
2. A measure of error

The confidence interval is computed by adding and subtracting the measure of error (and a standard value).

How do we compute a confidence interval

Two components:

1. A mean, coefficient, or mean difference
2. A measure of error

The confidence interval is computed by adding and subtracting the measure of error (and a standard value).

So, let's estimate heights of sixth grade students.

How do we compute a confidence interval

Two components:

1. A mean, coefficient, or mean difference
2. A measure of error

The confidence interval is computed by adding and subtracting the measure of error (and a standard value).

So, let's estimate heights of sixth grade students.

We observe a mean of 50in with a standard error of 5in. We would have the following CI:

How do we compute a confidence interval

Two components:

1. A mean, coefficient, or mean difference
2. A measure of error

The confidence interval is computed by adding and subtracting the measure of error (and a standard value).

So, let's estimate heights of sixth grade students.

We observe a mean of 50in with a standard error of 5in. We would have the following CI:

$$\text{Upper CI} = 50 + 1.96 * 5 = 59.8$$

$$\text{Lower CI} = 50 - 1.96 * 5 = 40.2$$

How do we compute a confidence interval

Two components:

1. A mean, coefficient, or mean difference
2. A measure of error

The confidence interval is computed by adding and subtracting the measure of error (and a standard value).

So, let's estimate heights of sixth grade students.

We observe a mean of 50in with a standard error of 5in. We would have the following CI:

$$\text{Upper CI} = 50 + 1.96 * 5 = 59.8$$

$$\text{Lower CI} = 50 - 1.96 * 5 = 40.2$$

We would write this as "95% CI [40.2, 59.8]"

Interpretation:

- We can be 95% confident that the true average height of students is between 40.2in and 59.8in.

Interpretation:

- We can be 95% confident that the true average height of students is between 40.2in and 59.8in.
- This is **not absolute certainty.**

Interpretation:

- We can be 95% confident that the true average height of students is between 40.2in and 59.8in.
- This is **not absolute certainty.**

A 95% confidence interval is a way to estimate the true value of a population parameter (what we are trying to estiamte from a sample).

Interpretation:

- We can be 95% confident that the true average height of students is between 40.2in and 59.8in.
- This is **not absolute certainty.**

A 95% confidence interval is a way to estimate the true value of a population parameter (what we are trying to estiamte from a sample).

It provides a range that likely includes the true value based on sample data.

Example 1

Testing the impact of a college degree on voting behavior.

Example 1

Testing the impact of a college degree on voting behavior.

Hypotheses:

Example 1

Testing the impact of a college degree on voting behavior.

Hypotheses:

H_a: Holding a college degree increase voter turnout.

Example 1

Testing the impact of a college degree on voting behavior.

Hypotheses:

H_a: Holding a college degree increase voter turnout. H_0: Holding a college degree has no effect on voter turnout.

Example 1

Testing the impact of a college degree on voting behavior.

Hypotheses:

H_a: Holding a college degree increase voter turnout. H_0: Holding a college degree has no effect on voter turnout.

Results:

Example 1

Testing the impact of a college degree on voting behavior.

Hypotheses:

H_a: Holding a college degree increase voter turnout. H_0: Holding a college degree has no effect on voter turnout.

Results:

Difference: 20% increase in voter turnout.

Example 1

Testing the impact of a college degree on voting behavior.

Hypotheses:

H_a: Holding a college degree increase voter turnout. H_0: Holding a college degree has no effect on voter turnout.

Results:

Difference: 20% increase in voter turnout. p-value: 0.04; t-statistic: 5.23; 95% Confidence Interval: [17.23, 22.14]

Example 1

Testing the impact of a college degree on voting behavior.

Hypotheses:

H_a: Holding a college degree increase voter turnout. H_0: Holding a college degree has no effect on voter turnout.

Results:

Difference: 20% increase in voter turnout. p-value: 0.04; t-statistic: 5.23; 95% Confidence Interval: [17.23, 22.14]

Conclusion:

Example 1

Testing the impact of a college degree on voting behavior.

Hypotheses:

H_a: Holding a college degree increase voter turnout. H_0: Holding a college degree has no effect on voter turnout.

Results:

Difference: 20% increase in voter turnout. p-value: 0.04; t-statistic: 5.23; 95% Confidence Interval: [17.23, 22.14]

Conclusion:

A college degree increases voting by 20% (95% CI [17.23, 22.14]). This is a statistically significant effect.

Example 2

Testing the relationship between anger and support for violence

Example 2

Testing the relationship between anger and support for violence

Hypotheses:

Example 2

Testing the relationship between anger and support for violence

Hypotheses:

H_a: Anger decreases support for violence.

Example 2

Testing the relationship between anger and support for violence

Hypotheses:

H_a: Anger decreases support for violence. H_0: Anger has no effect on support for violence.

Example 2

Testing the relationship between anger and support for violence

Hypotheses:

H_a: Anger decreases support for violence. H_0: Anger has no effect on support for violence.

- Control Group Support: 15%.

Example 2

Testing the relationship between anger and support for violence

Hypotheses:

H_a: Anger decreases support for violence. H_0: Anger has no effect on support for violence.

- Control Group Support: 15%.
- Angry Group Support: 72%.

Example 2

Testing the relationship between anger and support for violence

Hypotheses:

H_a: Anger decreases support for violence. H_0: Anger has no effect on support for violence.

- Control Group Support: 15%.
- Angry Group Support: 72%.
- Difference: 57%

Example 2

Testing the relationship between anger and support for violence

Hypotheses:

H_a: Anger decreases support for violence. H_0: Anger has no effect on support for violence.

- Control Group Support: 15%.
- Angry Group Support: 72%.
- Difference: 57%

P-Value $<$ 0.001; t-statistic: 19.23; 95% Confidence Interval: [53.7, 60.3]

Example 2

Testing the relationship between anger and support for violence

Hypotheses:

H_a: Anger decreases support for violence. H_0: Anger has no effect on support for violence.

- Control Group Support: 15%.
- Angry Group Support: 72%.
- Difference: 57%

P-Value $<$ 0.001; t-statistic: 19.23; 95% Confidence Interval: [53.7, 60.3]

Conclusion: Anger increases support for violence 15% (95% CI [53.7, 60.3]). This is a statistically significant effect.

Example 3

Testing the relationship between SAT scores and college graduation rates.

Example 3

Testing the relationship between SAT scores and college graduation rates.

Hypotheses:

Example 3

Testing the relationship between SAT scores and college graduation rates.

Hypotheses:

H_a: SAT scores increase college graduation rates.

Example 3

Testing the relationship between SAT scores and college graduation rates.

Hypotheses:

H_a: SAT scores increase college graduation rates. H_0: SAT scores have no effect on college graduation rates.

Example 3

Testing the relationship between SAT scores and college graduation rates.

Hypotheses:

H_a: SAT scores increase college graduation rates. H_0: SAT scores have no effect on college graduation rates.

- LOW SAT Group: 70%

Example 3

Testing the relationship between SAT scores and college graduation rates.

Hypotheses:

H_a: SAT scores increase college graduation rates. H_0: SAT scores have no effect on college graduation rates.

- LOW SAT Group: 70%
- High SAT Group: 73%

Example 3

Testing the relationship between SAT scores and college graduation rates.

Hypotheses:

H_a: SAT scores increase college graduation rates. H_0: SAT scores have no effect on college graduation rates.

- LOW SAT Group: 70%
- High SAT Group: 73%
- Difference: 3%

Example 3

Testing the relationship between SAT scores and college graduation rates.

Hypotheses:

H_a: SAT scores increase college graduation rates. H_0: SAT scores have no effect on college graduation rates.

- LOW SAT Group: 70%
- High SAT Group: 73%
- Difference: 3%

p-value: 0.04; t-statistic: 1.45; 95% Confidence Interval: [-1.2, 7.2]

Example 3

Testing the relationship between SAT scores and college graduation rates.

Hypotheses:

H_a: SAT scores increase college graduation rates. H_0: SAT scores have no effect on college graduation rates.

- LOW SAT Group: 70%
- High SAT Group: 73%
- Difference: 3%

p-value: 0.04; t-statistic: 1.45; 95% Confidence Interval: [-1.2, 7.2]

Conclusion: SAT scores are not significantly related to college graduation rates, with a difference in graduation rates of 3% (95% CI [-1.2, 7.2]). This is a statistically significant effect.

What kind of comparions might we want to test?

- Mean differences (ATEs)

What kind of comparions might we want to test?

- Mean differences (ATEs)
- Regression coefficients

What kind of comparions might we want to test?

- Mean differences (ATEs)
- Regression coefficients
- Differences in proportions (not covered in this class)

Differences in means

What is a t-test?

Differences in means

What is a t-test?

- A statistical test used to determine if there is a significant difference between the means of two groups.

Differences in means

What is a t-test?

- A statistical test used to determine if there is a significant difference between the means of two groups.
- Often used when sample sizes are small, and the population standard deviation is unknown.

Origins of the t-test

William Sealy Gosset (1876-1937)

Origins of the t-test

William Sealy Gosset (1876-1937)

- Developed the t-test in the early 1900s.

Origins of the t-test

William Sealy Gosset (1876-1937)

- Developed the t-test in the early 1900s.
- Worked as a chemist and statistician at the Guinness Brewery in Dublin, Ireland.

Origins of the t-test

William Sealy Gosset (1876-1937)

- Developed the t-test in the early 1900s.
- Worked as a chemist and statistician at the Guinness Brewery in Dublin, Ireland.
- Published under the pseudonym "Student," hence the name "Student's t-test."

The Brewery Challenge

- Problem at Guinness: Gosset needed a way to conduct small-sample experiments for quality control (e.g., barley quality, yeast consistency).

The Brewery Challenge

- Problem at Guinness: Gosset needed a way to conduct small-sample experiments for quality control (e.g., barley quality, yeast consistency).

Limitations:

The Brewery Challenge

- Problem at Guinness: Gosset needed a way to conduct small-sample experiments for quality control (e.g., barley quality, yeast consistency).

Limitations:

- Small sample sizes made it challenging to use standard statistical methods.

The Brewery Challenge

- Problem at Guinness: Gosset needed a way to conduct small-sample experiments for quality control (e.g., barley quality, yeast consistency).

Limitations:

- Small sample sizes made it challenging to use standard statistical methods.
- Large sample sizes were impractical due to costs and time.

The Brewery Challenge

- Problem at Guinness: Gosset needed a way to conduct small-sample experiments for quality control (e.g., barley quality, yeast consistency).

Limitations:

- Small sample sizes made it challenging to use standard statistical methods.
- Large sample sizes were impractical due to costs and time.

Limitations and Assumptions

Assumptions:

Limitations and Assumptions

Assumptions:

- Data should be approximately normally distributed (especially for small samples).

Limitations and Assumptions

Assumptions:

- Data should be approximately normally distributed (especially for small samples).
- Samples should have similar variances (especially for independent t-tests).

Limitations and Assumptions

Assumptions:

- Data should be approximately normally distributed (especially for small samples).
- Samples should have similar variances (especially for independent t-tests).

Limitations:

Limitations and Assumptions

Assumptions:

- Data should be approximately normally distributed (especially for small samples).
- Samples should have similar variances (especially for independent t-tests).

Limitations:

- Sensitive to outliers.

Limitations and Assumptions

Assumptions:

- Data should be approximately normally distributed (especially for small samples).
- Samples should have similar variances (especially for independent t-tests).

Limitations:

- Sensitive to outliers.
- May not be reliable if assumptions are significantly violated.

Limitations and Assumptions

Assumptions:

- Data should be approximately normally distributed (especially for small samples).
- Samples should have similar variances (especially for independent t-tests).

Limitations:

- Sensitive to outliers.
- May not be reliable if assumptions are significantly violated.