Kyiv
School of
Economics

Hypothesis testing

Ass. Professor Andriy Stavytskyy

Outline

- Hypothesis testing
- Example for MS Excel

Statistical Hypotheses

- A statistical test uses data from a sample to assess a claim about a population

- Statistical tests are framed formally in terms of two competing hypotheses:

- Null Hypothesis (H_0): Claim that there is no effect or difference.
- Alternative Hypothesis (H_a): Claim for which we seek evidence.

- Hypotheses are always about population parameters

Example: Sleep versus Caffeine – 1

- Students were given words to memorize, then randomly assigned to take either a 90 min nap, or a caffeine pill. 2 ½ hours later, they were tested on their recall ability.
- Explanatory variable: sleep or caffeine
- Response variable: number of words recalled
- Is sleep or caffeine better for memory?

Sleep versus Caffeine – 2

- Let μ_s and μ_c be the mean number of words recalled after sleeping and after caffeine.
- Is there a difference in average word recall between sleep and caffeine?
- What are the null and alternative hypotheses?

- H_0: $\mu_s \neq \mu_c$, H_a: $\mu_s = \mu_c$
- H_0: $\mu_s = \mu_c$, H_a: $\mu_s \neq \mu_c$
- H_0: $\mu_s \neq \mu_c$, H_a: $\mu_s > \mu_c$
- H_0: $\mu_s = \mu_c$, H_a: $\mu_s > \mu_c$
- H_0: $\mu_s = \mu_c$, H_a: $\mu_s < \mu_c$

- The null hypotheses is “no difference,” or that the means are equal. The alternative hypothesis is that there is a difference.

Exercise

- Take a minute to write down the hypotheses for each of the following situations:
  - Does the proportion of people who support gun control differ between males and females?
  - Is the average hours of sleep per night for college students less than 7?

Solution

- Does the proportion of people who support gun control differ between males and females?

p_f: proportion of females who support gun control
p_m: proportion of males who support gun control

$H_0: p_f = p_m$
$H_a: p_f \neq p_m$

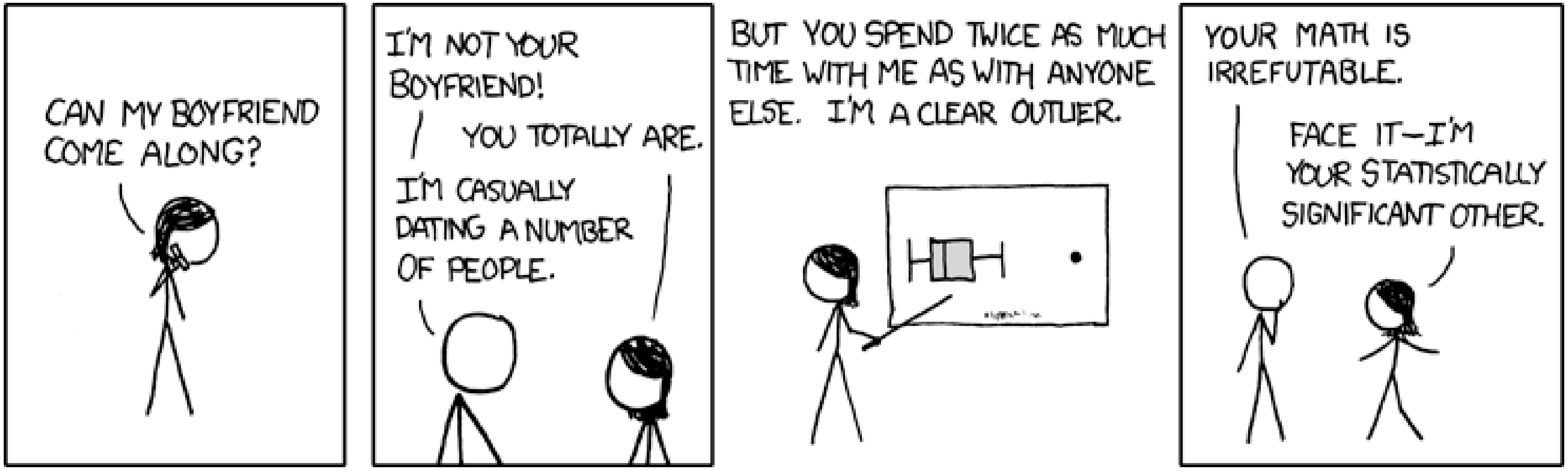
- Is the average hours of sleep per night for college students less than 7?

μ: average hours of sleep per night for college students

$H_0: \mu = 7$
$H_a: \mu < 7$

Statistical Significance

- When results as extreme as the observed sample statistic are unlikely to occur by random chance alone (assuming the null hypothesis is true), we say the sample results are statistically significant
- If our sample is statistically significant, we have convincing evidence against H_0, in favor of H_a
- If our sample is not statistically significant, our test is inconclusive



Sleep versus Caffeine

- μ_s and μ_c: mean number of words recalled after sleeping and after caffeine
- H_0: $\mu_s = \mu_c$, H_a: $\mu_s \neq \mu_c$
- The sample difference in means is $\bar{x}_s - \bar{x}_c = 3$, and this is statistically significant. We can conclude…

- there is a difference between sleep and caffeine for memory (and data show sleep is better)
- there is not a difference between sleep and caffeine for memory
- nothing

Short summary

- Statistical tests use data from a sample to assess a claim about a population
- Statistical tests are usually formalized with competing hypotheses:
  - Null hypothesis (H_0): no effect or no difference
  - Alternative hypothesis (H_a): what we seek evidence for
- If data are statistically significant, we have convincing evidence against the null hypothesis, and in favor of the alternative

Hypothesis testing

- The process of making judgments about a large group (population) on the basis of a small subset of that group (sample) is known as **statistical inference**.
- Hypothesis testing, one of two fields in statistical inference, allows us to objectively assess the probability that statements about a population are true. Because these statements are probabilistic in nature, we **can never be certain of their truth**.
- Steps in hypothesis testing
  1. Stating the hypotheses.
  2. Identifying the appropriate test statistic and its probability distribution.
  3. Specifying the significance level.
  4. Stating the decision rule.
  5. Collecting the data and calculating the test statistic.
  6. Making the statistical decision.
  7. Making the economic or investment decision.

1. State the hypothesis – 1

- The foundation of hypothesis testing lies in determining exactly what we are to test.
- We organize a hypothesis test into two categories.
  - The null hypothesis, denoted H_0, is the hypothesis we are testing.
  - The alternative hypothesis is denoted H_a.
- The different possibilities represented by the two hypotheses should be mutually exclusive and collectively exhaustive.

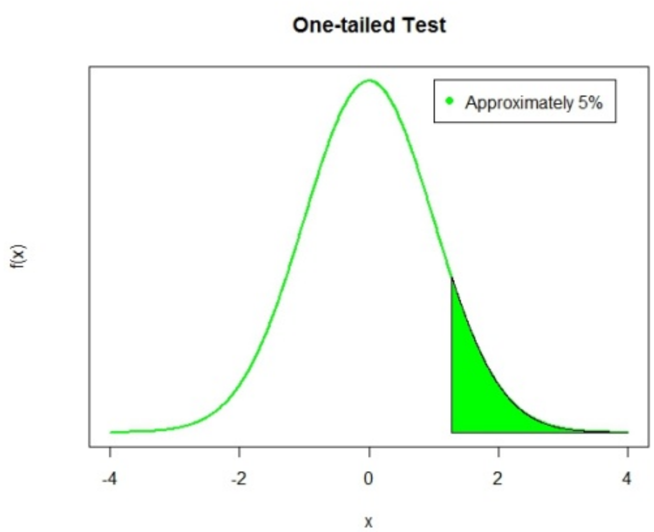
- Three different ways of formulating a hypothesis test:
  - H_0: $\theta = \theta_0$ versus H_a: $\theta \neq \theta_0$ (a “not equal to” alternative hypothesis)
  - H_0: $\theta \leq \theta_0$ versus H_a: $\theta > \theta_0$ (a “greater than” alternative hypothesis)
  - H_0: $\theta \geq \theta_0$ versus H_a: $\theta < \theta_0$ (a “less than” alternative hypothesis)
- Hypothesis tests generally concern the true value of a population parameter as determined using a sample statistic.

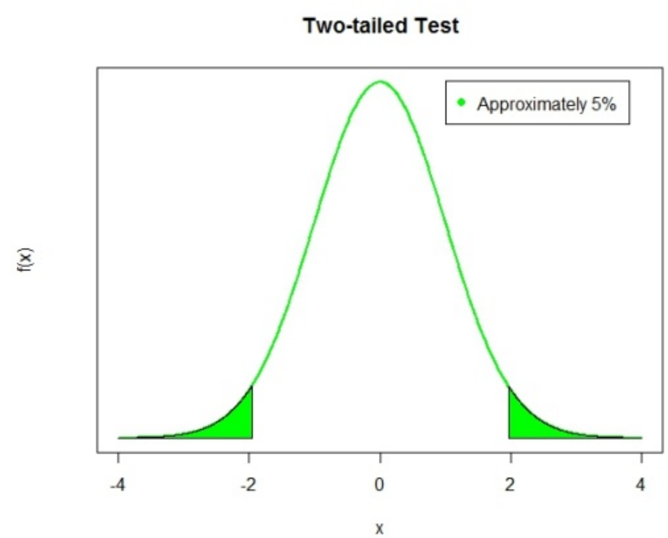
1. State the hypothesis – 2

- The hypothesis is designed to assess the likelihood of a sample statistic accurately representing the population statistic it attempts to measure.
- Hypothesis tests are formulated in such a way that they lead to either one-tailed tests or two-tailed tests.
- One-tailed tests are comparisons based on a single side of the distribution, whereas two-tailed tests admit the possibility of the true population parameter lying in either tail of the distribution.
  - H_0: $\theta = \theta_0$ versus H_a: $\theta \neq \theta_0$ (a two-tailed test)
  - H_0: $\theta \leq \theta_0$ versus H_a: $\theta > \theta_0$ (a one-tailed test for the upper tail)
  - H_0: $\theta \geq \theta_0$ versus H_a: $\theta < \theta_0$ (a one-tailed test for the lower tail)

1. State the hypothesis – 3

- The selection of an appropriate null hypothesis and, as a result, an alternative hypothesis, center around economic or financial theory as it relates to the point estimate(s) being tested.
  - Two-tailed tests are more “conservative” than one-tailed tests. In other words, they lead to a fail-to-reject the null hypothesis conclusion more often.
  - One-tailed tests are often used when financial or economic theory proposes a relationship of a specific direction.





2. Identifying the appropriate test statistic and its probability distribution

- The test statistic is a measure based on the difference between the hypothesized parameter and the sample point estimate that is used to assess the likelihood of that sample statistic resulting from the underlying population.
- Test statistics that we implement will generally follow one of the following distributions:
  - t-distribution
  - Standard normal z
  - F-distribution
  - Chi-square distribution

Errors in Hypothesis tests

- Type I errors occur when we reject a null hypothesis that is actually true. Type II errors occur when we do not reject a null hypothesis that is false.

Decision	True Situation	
	H_0: True	H_0: False
Do not reject	Correct decision	Type II error
Reject	Type I error	Correct decision*

- Mutually exclusive problems:
  - If we mistakenly reject the null, we make a Type I error.
  - If we mistakenly fail to reject the null, we make a Type II error.
  - Because we can't reject and fail to reject simultaneously because of the mutually exclusive nature of the null and alternative hypothesis, the errors are also mutually exclusive.

* The rate at which we correctly reject a false null hypothesis is known as the power of the test.

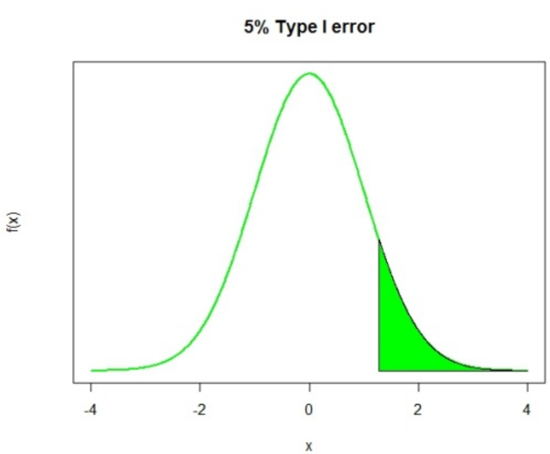
3. Specifying the significance level

- The level of significance is the desired standard of proof against which we measure the evidence contained in the test statistic.
- The level of significance is identical to the level of a Type I error and, like the level of a Type I error, is often referred to as “alpha,” or a.
- The level of confidence in the statistical results is directly related to the significance level of the test and, thus, to the probability of a Type I error.

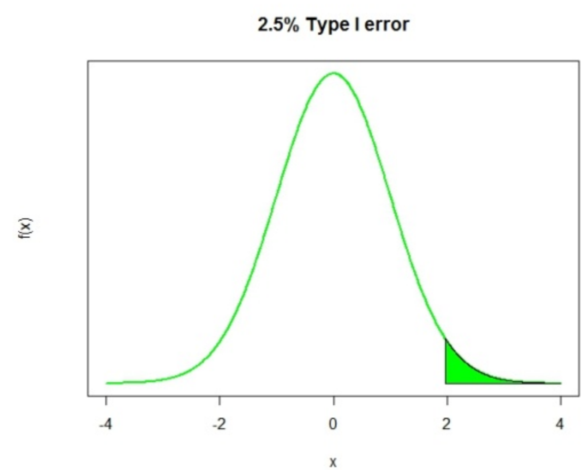
Significance Level	Suggested Description
0.10	“some evidence”
0.05	“strong evidence”
0.01	“very strong evidence”

The Trade-off in hypothesis testing

- Because the significance level and the Type I error rate are the same, and Type I and Type II rates are mutually exclusive, there is a trade-off in setting the significance level.
  - If we decrease the probability of a Type I error by specifying a smaller significance level, we increase the probability of a Type II error.
  - The only way to decrease the probability of both errors at the same time is to increase the sample size because such an increase reduces the denominator of our test statistic.

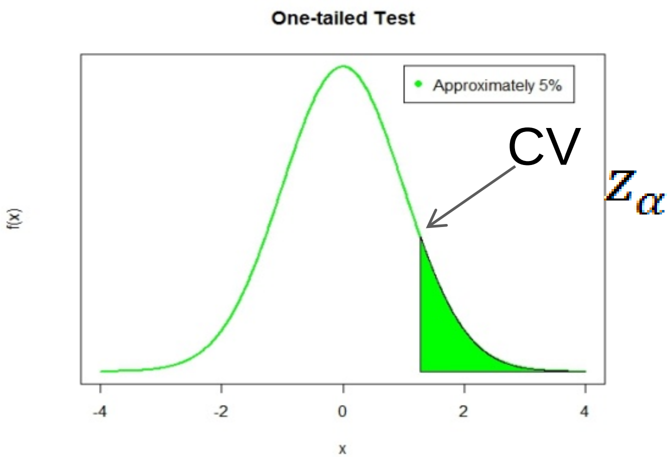


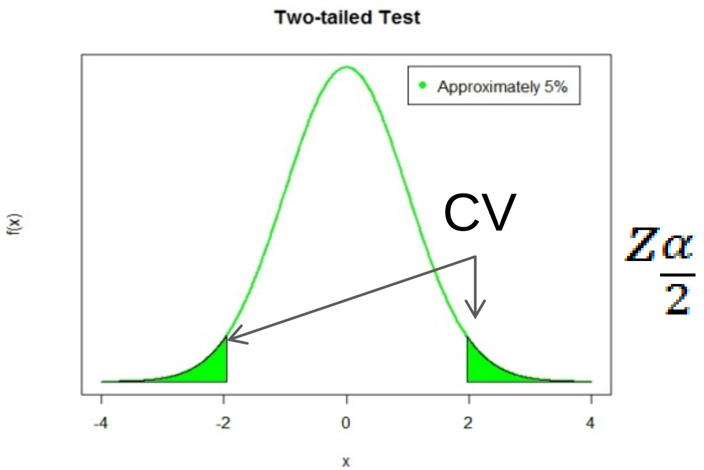
Decreased Type I
but
Increased Type II



4. Stating the decision rule

- The decision rule uses the significance level and the probability distribution of the test statistic to determine the value above (below) which the null hypothesis is rejected.
- The critical value (CV) of the test statistic is the value above (below) which the null hypothesis is rejected.
  - Also known as a rejection point.
  - One-tailed tests are indicated with a subscript α.
  - Two-tailed tests are indicated with a subscript α/2.





The p-value approach

- The p-value is the smallest level of significance at which a given null hypothesis can be rejected.
- The selection of a particular level of significance is somewhat arbitrary.
  - Lower levels lead to greater confidence but come at an increased risk of Type II errors.

Confidence Interval or Hypothesis test?

- Two-tailed hypothesis tests can easily be rewritten as confidence intervals.
  - Recall that a two-tailed hypothesis test rejects the null when the observed value of the test statistic is either below the lower critical value or above the upper.
  - The lower critical value can be restated as the lower limit on a confidence interval.
  - The upper critical value can be restated as the upper limit on a confidence interval.

$$[\bar{X} - z_{\frac{\alpha}{2}}(\sigma_{\bar{X}}), \bar{X} + z_{\frac{\alpha}{2}}(\sigma_{\bar{X}})]$$

  - When the hypothesized population parameter lies within this confidence interval, we fail to reject the null hypothesis.
  - Although this relationship is useful, it precludes easy calculation of the significance level of the test, known as a p-value, from the values of the standard error and point estimate.

5. Collect the data and calculate the test statistic

- In practice, data collection is likely to represent the largest portion of the time spent in hypothesis testing, and care should be given to the sampling considerations, particularly biases introduced in the data collection process.

6. Make the statistical decision

- The statistical process is completed when we compare the test statistic from Step 5 with the critical value in Step 4 and assess the statistical significance of the result.
- Reject or fail to reject the null hypothesis.

7. Make the economic decision

- Quantitative analysis is used to guide decision making in a scientific manner; hence, the end of the process lies in making a decision.
- The economic or investment decision should take into account not only the statistical evidence, but also the economic value of acting on the statistical conclusion.
  - We may find strong statistical evidence of a difference but only weak economic benefit to acting.
  - Because the statistical process often focuses only on one attribute of the data, other attributes may affect the economic value of acting on our statistical evidence.
    - For example, a statistically significant difference in mean return for two alternative investment strategies may not lead to economic gain if the higher-returning strategy has much higher transaction costs.
- The economic forces leading to the statistical outcome should be well understood before investing.

Possible hypothesis

Testing a single mean

- H_0: $\mu=\mu_0$
- H_1: $\mu\neq\mu_0$

Testing for a difference in means

- H_0: $\mu_1-\mu_2=\mu_0$
- H_1: $\mu_1-\mu_2\neq\mu_0$

Testing a single variance

- H_0: $\sigma^2=\sigma_0$
- H_1: $\sigma^2\neq\sigma_0$

Testing for equality of variance

- H_0: $\sigma_1^2=\sigma_2^2$
- H_1: $\sigma_1^2\neq \sigma_2^2$

Testing a single mean – 1

- H_0: $\mu=\mu_0$
- H_1: $\mu\neq\mu_0$

- We almost never know the variance of the underlying population, and in such cases, tests of a single mean are either t-tests or z-tests.

Testing a single mean – 2

- Tests comparing a single mean with a value:
  - Use a t-test with df = n – 1 when
    - Population variance is unknown and
    - Sample is large or sample is small but (approximately) normally distributed.
    - $t_{n-1} = \frac{\bar{X} - \mu_0}{\hat{s}/\sqrt{n}}$
  - Use a z-test if
    - the sample is large or
    - the population is normally distributed.
    - $z = \frac{\bar{X} - \mu_0}{\hat{s}/\sqrt{n}}$

- Note that two of these use the sample standard deviation as an estimate of population standard deviation.

Testing a single mean: Example – 1

- You have collected data on monthly equity returns and determined that the average return across the 48-month period you are examining was 12.94% with a standard deviation of returns of 15.21%. You want to test whether this average return is equal to the 15% return that your retirement models use as an underlying assumption. You want to be 95% confident of your results.

Testing a single mean: Example – 2

- Formulate hypothesis
  H_0: θ = 15%
  H_a: θ ≠ 15% (a two-tailed test).
- Identify appropriate test statistic
  t-test for an unknown population variance.
- Specify the significance level
  0.05 as stated in the setup
- Calculate a critical value
  t=2.01174.
- Collect data (see above) and calculate test statistic

$$t_{n-1} = \frac{\bar{X} - \mu_0}{\hat{s}/\sqrt{n}}, \qquad t_{48-1} = \frac{0.1294 - 0.15}{0.1521/\sqrt{48}} = -0.938$$

- Make the statistical decision
  - Do not reject the null hypothesis.

Testing for a difference in means: Independent Samples

- H_0: $\mu_1 - \mu_2 = \mu_0$
- H_1: $\mu_1 - \mu_2 \neq \mu_0$

- Normally distributed, equal but unknown variances
  - Uses a pooled variance estimator, s_p^2, which is a weighted average of the sample variances.

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - \mu_0}{\sqrt{\left(\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}\right)}} \qquad s_p^2 = \frac{(n_1 - 1)\hat{s}_1^2 + (n_2 - 1)\hat{s}_2^2}{n_1 + n_2 - 2} \qquad \mathrm{df} = n_1 + n_2 - 2$$

- Normally distributed, unequal and unknown variances
  - Uses a different pooled variance estimator

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - \mu_0}{\sqrt{\left(\frac{\hat{s}_1^2}{n_1} + \frac{\hat{s}_2^2}{n_2}\right)}} \qquad \mathrm{df} = n_1 + n_2 - 2$$

Testing for a difference in means: Example – 1

- You have decided to investigate whether the return to your client's retirement portfolio will be enhanced by the addition of foreign equities. Accordingly, you first want to test whether foreign equities have the same return as domestic equities before proceeding with further analysis.
- U.S. equities returned 12.94% with a standard deviation of 15.21% over the prior 48 months. You have determined that foreign equities returned 17.67% with a standard deviation of 16.08% over the same period. You want the same level of confidence in this result (5%).
- You are willing to assume, for now, that the two samples are independent, approximately normally distributed, and drawn from a population with the same underlying variance.

Testing for a difference in means: Example – 2

- Stating the hypotheses

  H_0: $\mu_{DomEq} - \mu_{ForEq} = 0$

  H_a: $\mu_{DomEq} - \mu_{ForEq} \neq 0$

- Identifying the appropriate test statistic and its probability distribution
  - t-test for unequal means with a normal distribution and unknown but equal variances
- Specifying the significance level and critical value
  - $\alpha=0.05$, CV = –1.986
- Stating the decision rule
  - Reject the null if |TS| > 1.986

Testing for a difference in means: Example – 3

- Collecting the data and calculating the test statistic →

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - \mu}{\sqrt{\left(\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}\right)}}$$

$$s_p^2 = \frac{(n_1 - 1)\hat{s}_1^2 + (n_2 - 1)\hat{s}_2^2}{n_1 + n_2 - 2}$$

$$s_p^2 = \frac{(47)0.1521^2 + (47)0.1608^2}{48 + 48 - 2}$$

$$t = \frac{(0.1294 - 0.1767)}{\sqrt{\left(\frac{0.0242^2}{48} + \frac{0.0242^2}{48}\right)}} = -1.4806$$

$$\text{df} = 48 + 48 - 2$$

- Making the statistical decision → fail to reject

Testing a single variance

- H_0: $\sigma^2 = \sigma_0$
- H_1: $\sigma^2 \neq \sigma_0$

- Tests of a single variance
  - Normally distributed population
  - Chi-square test with df = n – 1
  - Very sensitive to underlying assumptions
  - Test statistic

$$\chi^2_{n-1} = \frac{(n-1)\hat{s}^2}{\sigma_0^2}$$

Testing a single variance: Example

- Is the variance of domestic equity returns from our previous example, 15.21%, statistically different from 10%?

$$\chi^2 = \frac{(n-1)\hat{s}^2}{\sigma_0^2} \qquad \chi^2 = \frac{47 \cdot 15.21^2}{10^2} = 108.73$$

- Critical value for α = 5% is 64.0011 → Reject the null

Testing for equality of variance

- H_0: $\sigma_1^2 = \sigma_2^2$
- H_1: $\sigma_1^2 \neq \sigma_2^2$

Tests comparing two variance measures:

- If we have two normally distributed populations, then a ratio test of the two variances follows an F-distribution.

$$F(\mathrm{df}_1, \mathrm{df}_2) = \frac{\hat{s}_1^2}{\hat{s}_2^2}$$

$$\mathrm{df}_i = n_i - 1$$

- If the test statistic is greater than the critical value for an F-distribution with df_1 and df_2 degrees of freedom, reject the null.

Testing for equality of variance: Example

- Return now to our earlier example comparing foreign and domestic equity returns. In the example, we assumed that the variances were equal. Perform the necessary test to assess the validity of this assumption. We had 48 observations for each return series, foreign equity returns had a standard deviation of 16.08%, and domestic of 15.21%.

Testing for equality of variance

- Stating the hypotheses
  H_0: $s_{DomEq} = s_{ForEq}$
  H_a: $s_{DomEq} \neq s_{ForEq}$
- Identifying the appropriate test statistic and its probability distribution
  F-test for a ratio of variances
- Specifying the significance level
  CV = 1.6238
- Stating the decision rule
  Reject the null if TS > 1.6238
- Collecting the data and calculating the
  - test statistic →

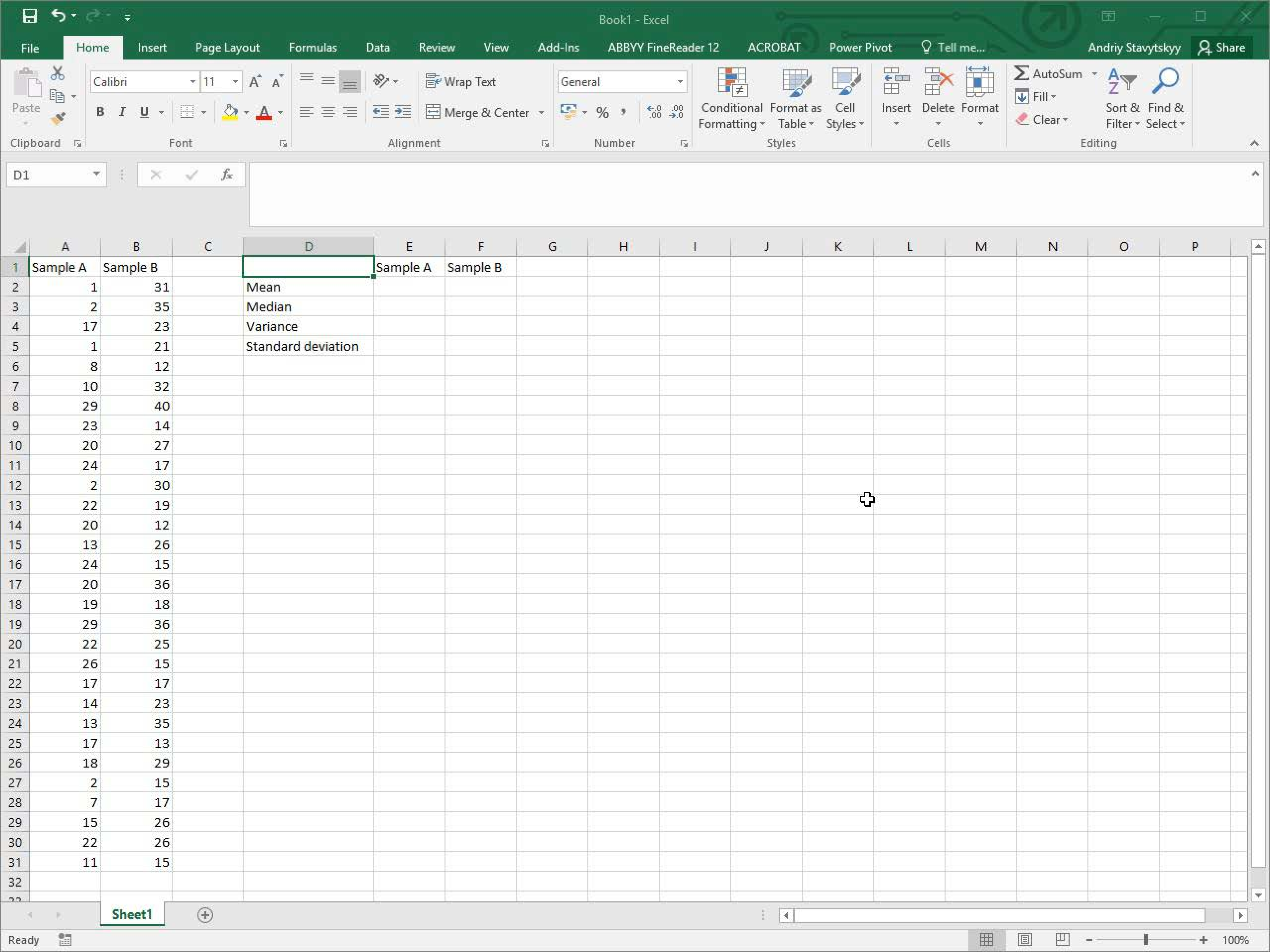
$$F(\mathrm{df}_1, \mathrm{df}_2) = \frac{\hat{s}_1^2}{\hat{s}_2^2} \qquad \mathrm{df}_i = n_i - 1 \qquad F(47, 47) = \frac{0.1608^2}{0.1521^2} = 1.1177$$

- Making the statistical decision → Fail to reject

Example for MS Excel

- Provide descriptive statistics for two samples
- Test hypothesis

- See video



Book1 - Excel
File
Home
Insert
Page Layout
Formulas
Data
Review
View
Add-Ins
ABBYY FineReader 12
ACROBAT
Power Pivot
Tell me...
Andriy Stavytskyy
Share
Paste
Calibri
11
Wrap Text
Merge & Center
General
Conditional Formatting
Format as Table
Cell Styles
Insert
Delete
Format
AutoSum
Fill
Clear
Sort & Filter
Find & Select
Clipboard
Font
Alignment
Number
Styles
Cells
Editing
D1
Sample A
Sample B
1 31
2 35
17 23
1 21
8 12
10 32
29 40
23 14
20 27
24 17
2 30
22 19
20 12
13 26
24 15
20 36
19 18
29 36
22 25
26 15
17 17
14 23
13 35
17 13
18 29
2 15
7 17
15 26
22 26
11 15
Sample A
Sample B
Mean
Median
Variance
Standard deviation
Sheet1
Ready
100%

REVIEW

Short summary

- Statistical tests use data from a sample to assess a claim about a population
- Statistical tests are usually formalized with competing hypotheses:
  - Null hypothesis (H_0): no effect or no difference
  - Alternative hypothesis (H_a): what we seek evidence for
- If data are statistically significant, we have convincing evidence against the null hypothesis, and in favor of the alternative

Hypothesis testing

- The process of making judgments about a large group (population) on the basis of a small subset of that group (sample) is known as **statistical inference**.
- Hypothesis testing, one of two fields in statistical inference, allows us to objectively assess the probability that statements about a population are true. Because these statements are probabilistic in nature, we **can never be certain of their truth**.
- Steps in hypothesis testing
  1. Stating the hypotheses.
  2. Identifying the appropriate test statistic and its probability distribution.
  3. Specifying the significance level.
  4. Stating the decision rule.
  5. Collecting the data and calculating the test statistic.
  6. Making the statistical decision.
  7. Making the economic or investment decision.

Possible hypothesis

Testing a single mean

- H_0: $\mu=\mu_0$
- H_1: $\mu\neq\mu_0$

Testing for a difference in means

- H_0: $\mu_1-\mu_2=\mu_0$
- H_1: $\mu_1-\mu_2\neq\mu_0$

Testing a single variance

- H_0: $\sigma^2=\sigma_0$
- H_1: $\sigma^2\neq\sigma_0$

Testing for equality of variance

- H_0: $\sigma_1^2=\sigma_2^2$
- H_1: $\sigma_1^2\neq \sigma_2^2$

Thank you for attention!

KSE