ISE 315: Engineering Statistics

Lecture 16: Software Tools for Statistical Inference

Instructor: Mansur M. Arief, PhD
Industrial and Systems Engineering, KFUPM

Office: 22-219 — Email: mansur.arief@kfupm.edu.sa

Based on Montgomery & Runger, Applied Statistics and Probability for Engineers, 6th Ed.

Lecture 16

Software Tools for Automating Statistical Inference & Regression

Lecture Outline

- Why use software for statistical inference?
- Overview of available tools: Minitab, Excel, Python, R
- Minitab: confidence intervals, hypothesis tests, regression
- What the software automates vs. what **you** still need to do
- Demo: Python/Jupyter notebook for Chapters 8–10

Why Use Software?

From tables to tools

What we have been doing:

- Computing test statistics by hand: Z_0, T_0, χ^2_0, F_0
- Looking up critical values from appendix tables
- Bounding p-values between table entries

What software gives us:

- Exact p-values (e.g., $p = 0.0297$ instead of $0.02 < p < 0.05$)
- Graphical diagnostics (residual plots, distribution checks)
- Handles large datasets and complex designs effortlessly

Software speeds up *calculation*, but still need to understand the results.

Statistical Software Landscape

Tool	Strengths	Limitations	Typical Use
Minitab	Menu-driven, built for quality/industrial stats	Commercial license, less flexible	Six Sigma, QC, DoE in industry
Excel	Widely available, familiar interface	Limited stat functions, rounding issues	Quick calculations, data entry
Python (`scipy`)	Free, fully programmable, reproducible	Requires coding knowledge	Research, automation, ML
R	Purpose-built for statistics, rich packages	Steeper learning curve	Academic research, biostatistics

In ISE 315: We focus on understanding the methods. Any of these tools can execute them. Today we look at Minitab (industry standard for ISE) and Python (free and programmable).

Minitab

Walkthrough for Chapters 8–10

Minitab: Confidence Interval on the Mean

What you provide: Data column **or** summary stats ($n, \bar{x}, s$ or σ), confidence level.

Minitab — 1-Sample Z: WallThickness

Descriptive Statistics

N	Mean	StDev	SE Mean	95% CI for μ
25	0.27310	0.02000	0.00400	(0.26527, 0.28093)

μ: population mean of WallThickness

Known standard deviation = 0.02

Verify by hand: $\bar{x} \pm z_{0.025}\frac{\sigma}{\sqrt{n}} = 0.2731 \pm 1.96 \times 0.004 = (0.2653, 0.2809)$ ✓

Minitab: CI When σ Is Unknown

Scenario: $n = 25$ energy cost measurements. We do not know σ.

Minitab — 1-Sample t: EnergyCost

```
Descriptive Statistics
 N      Mean     StDev    SE Mean    95% CI for μ
25    330.60    154.20      30.84    (266.94, 394.26)

μ:  population mean of EnergyCost
```

Note the differences from the Z-interval:

- Uses $s = 154.20$ (sample StDev) instead of known σ
- Uses $t_{0.025,\,24} = 2.064$ instead of $z_{0.025} = 1.96$
- Resulting CI is **wider** due to extra uncertainty in estimating σ

Minitab: Hypothesis Test on the Mean

Chemical yield, $n = 10$, $\bar{x} = 82.5$, $s = 3.2$. Test $H_0 : \mu = 80$ at $\alpha = 0.05$.

Minitab — 1-Sample t: Yield

Descriptive Statistics

N	Mean	StDev	SE Mean	95% CI for μ
10	82.500	3.200	1.012	(80.212, 84.788)

μ: population mean of Yield

Test

Null hypothesis H_0: $\mu = 80$
Alternative hypothesis H_1: $\mu \neq 80$

T-Value	P-Value
2.47	0.035

By hand: $T_0 = 2.471$, from the t-table $0.02 < \text{p-value} < 0.05$.
Minitab gives the **exact** p-value: $p = 0.035$. **Reject** H_0.

Minitab: Test on Variance

Pipeline wall thickness, $n = 20$, $s = 0.018$ in. Test $H_0 : \sigma^2 = 0.0004$ at $\alpha = 0.05$.

Minitab — 1 Variance: WallThickness

```
Test
Null hypothesis          H0:  σ² = 0.0004
Alternative hypothesis   H1:  σ² ≠ 0.0004

Method:  Chi-Square

Statistic   DF   P-Value
    15.39   19     0.697

95% CI for σ²:  (0.000162, 0.000717)
```

By hand: $\chi^2_0 = \frac{(20-1)(0.018)^2}{0.0004} = \frac{19 \times 0.000324}{0.0004} = 15.39$
From the χ^2-table, $p > 0.10$. Minitab gives $p = 0.697$: **fail to reject H_0**.

Minitab also reports CI for σ^2 and σ, plus Bonett method for non-normal data.

Minitab: Two-Sample t-Test

Tensile strength from two production lines. Test $H_0 : \mu_1 = \mu_2$ at $\alpha = 0.05$.

Minitab — 2-Sample t: Line1 vs Line2 (pooled)

Descriptive Statistics

Sample	N	Mean	StDev	SE Mean
Line1	10	48.50	2.30	0.73
Line2	12	47.10	2.50	0.72

Difference	$\mu_1 - \mu_2$
Estimate for difference	1.400
95% CI for difference	$(-0.334, 3.134)$

Test

Null hypothesis	$H_0: \mu_1 - \mu_2 = 0$
Alternative hypothesis	$H_1: \mu_1 - \mu_2 \neq 0$

T-Value	DF	P-Value
1.68	20	0.108

Both use Pooled StDev = 2.4166

Minitab: Complete Two-Sample Menu Map

Test	Minitab Menu Path	Key Options
Two-sample t (pooled)	`Stat` → `Basic Stat` → `2-Sample t`	Check "Assume equal variances"
Two-sample t (Welch)	`Stat` → `Basic Stat` → `2-Sample t`	Uncheck equal variances
Paired t	`Stat` → `Basic Stat` → `Paired t`	Two matched columns
Two variances (F-test)	`Stat` → `Basic Stat` → `2 Variances`	Levene's test also shown
Two proportions	`Stat` → `Basic Stat` → `2 Proportions`	Event counts and trials

Minitab automatically provides:

- Both the test statistic, exact p-value, and confidence interval
- Degrees of freedom (including Satterthwaite's for Welch)

Tip: Run `2 Variances` first. If the F-test rejects equal variances, uncheck "Assume equal variances" in the 2-Sample t dialog.

Minitab: Regression Analysis

Chapters 11–12 Preview — Stat > Regression > Fit Regression Model

Minitab — Regression: Strength vs Temperature

Coefficients

Term	Coef	SE Coef	T-Value	P-Value
Constant	127.80	5.22	24.49	0.000
Temperature	−0.584	0.061	−9.57	0.000

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
5.124	83.60%	82.50%	79.83%

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	2403.5	2403.5	91.58	0.000
Error	18	472.7	26.3		
Total	19	2876.2			

What Minitab gives you here: $\hat{\beta}_0, \hat{\beta}_1$ with t-tests, R^2, full ANOVA table, and the F-test for overall significance — all in one output.

Minitab: Residual Diagnostics

The "Four-in-One" Plot

Minitab's `Graphs` button in the regression menu gives **four-in-one residual plot**:

Normal Probability Plot of Residuals (checks normality)	**Versus Fits** Residuals vs. $\hat{y}$ (checks constant variance)
Histogram of Residuals (checks distribution shape)	**Versus Order** Residuals vs. run order (checks independence)

What Software Automates

Software handles this	You still must do this
Compute test statistics (Z_0, T_0, χ^2_0, F_0)	Choose the correct test for the problem
Look up exact p-values	Interpret the p-value in context
Calculate CI bounds	Select the right confidence level
Degrees of freedom (incl. Welch)	Verify assumptions (normality, independence)
Residual plots	Judge whether assumptions are met
Regression coefficients	Decide which variables to include
ANOVA decomposition	Explain what the results mean to a stakeholder

Warning: Software will happily run a t-test on non-normal data, or fit a regression with meaningless predictors. It does not check whether your analysis makes sense.

What We Still Need From You

1. **Problem Formulation**
   - Writing H_0 and H_1 correctly (one-sided vs. two-sided)
   - Choosing α based on the engineering context and consequences
2. **Test Selection**
   - σ known $\rightarrow Z$, unknown $\rightarrow t$, variance $\rightarrow \chi^2$, two variances $\rightarrow F$
3. **Assumption Checking**
   - Is the population approximately normal? Is n large enough for CLT?
   - Are the two samples truly independent?
   - For proportions: are $n\hat{p} \geq 5$ and $n(1 - \hat{p}) \geq 5$?
4. **Interpretation and Communication**

Suggested Workflow

Problem: A refinery claims their catalyst reduces SO_2 emissions. An engineer collects $n = 15$ measurements before and after. Is the mean emission lower after using the catalyst?

You (the engineer) decide:

- This is **paired data** (same equipment, before vs. after)
- $H_0 : \mu_D = 0$ vs. $H_1 : \mu_D > 0$ (where $D_i = \text{Before}_i - \text{After}_i$)
- Use $\alpha = 0.05$

Software computes: $\bar{d} = 3.2$, $s_D = 4.8$, $T_0 = 2.582$, $p = 0.011$

You (the engineer) conclude: "At the 5% significance level, there is sufficient evidence that the catalyst reduces mean SO_2 emissions."

Python Alternative

Python for Statistical Inference

Key functions in `scipy.stats`:

Our Test	Python Function	Returns
1-sample Z (CI/HT)	`norm.cdf()`, `norm.ppf()`	CDF values, critical values
1-sample t-test	`ttest_1samp(data, μ0)`	T_0, p-value
2-sample t-test	`ttest_ind(data1, data2)`	T_0, p-value
Paired t-test	`ttest_rel(before, after)`	T_0, p-value
χ^2 on variance	`chi2.cdf()`, `chi2.ppf()`	CDF values, critical values
F-test on variances	`f.cdf()`, `f.ppf()`	CDF values, critical values
Proportion Z-test	`proportions_ztest()`	Z_0, p-value

Advantage: You can write a script once and rerun it on any new dataset. Fully reproducible, free, and auditable.

Python Example: One-Sample *t*-Test

```python
from scipy import stats
import numpy as np

data = [82.1, 84.3, 79.5, 83.8, 81.2, 85.1, 80.9, 83.4, 82.7, 81.5]

# H0: mu = 80 vs H1: mu != 80, alpha = 0.05
mu_0 = 80
t_stat, p_value = stats.ttest_1samp(data, mu_0)

print(f"T-statistic: {t_stat:.4f}")
print(f"P-value:     {p_value:.4f}")

# Confidence interval
n = len(data)
xbar, s = np.mean(data), np.std(data, ddof=1)
t_crit = stats.t.ppf(0.975, df=n-1)
ci = (xbar - t_crit*s/np.sqrt(n),
      xbar + t_crit*s/np.sqrt(n))
print(f"95% CI: ({ci[0]:.3f}, {ci[1]:.3f})")
```

Common Pitfalls When Using Software

- **Blindly trusting the output.**
  Software does not check your assumptions. A t-test on highly skewed data with $n = 5$ will still produce a p-value, but it may not be valid.
- **Choosing the wrong test.**
  Running a two-sample t-test on paired data inflates the standard error and loses power. The software will not warn you.
- **Ignoring practical significance.**
  With $n = 10{,}000$, even a trivial difference can be "statistically significant" ($p < 0.001$). Always report effect size alongside the p-value.
- **Forgetting one-sided vs. two-sided.**
  Most software defaults to a two-sided test. If your H_1 is one-sided, you may need to divide the reported p-value by 2, or select the correct option.

Demo: ISE 315 Statistical Inference Notebook

We have prepared a **Jupyter Notebook** that covers all tests from Chapters 8–10:

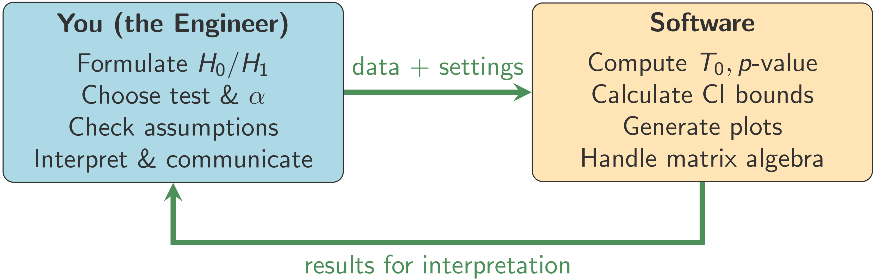
Features:

- Upload your raw data (CSV) or enter summary statistics
- Select the test type from a menu
- The notebook recommends settings based on your inputs
- Outputs: test statistic, exact p-value, CI, and a plain-English conclusion

Supported tests:

- Chapter 8: CI for μ (σ known/unknown), CI for σ^2, CI for p
- Chapter 9: Z-test, t-test, χ^2-test on variance, proportion test
- Chapter 10: Two-sample Z, pooled t, Welch t, paired t, F-test, two-proportion Z

Summary: The Right Balance



Bottom line: Master the concepts first (that is what exams test). Then use software to be efficient and accurate in practice.

Remarks

- The Jupyter notebook is posted on Blackboard (Supplementary Materials)
- Software is useful for homework verification and your own projects
- See you after the break! Ramadan Mubarak!