1.0 Hypothesis Testing

The Six Steps of Hypothesis Testing
Rejecting or Accepting the Null
Samples and Data

1.1 Parametric Tests for Comparing Two Populations

Testing Equality of Population Means (T Test)

$$\left[\frac{\bar{X} \pm t_{0.025,n-1}s}{\sqrt{n}}\right]$$

Independent Samples with Unequal Population Variances
Independent Samples and Equal Population Variances
Dependent Samples – Matched Pairs
Testing Equality of Population Variances (F Distribution or F Test)

$$\left[\frac{(n-1)s^2}{\chi^2_{a\backslash 2,,n-1}}, \frac{(n-1)s^2}{\chi^2_{1-\alpha,n-1}}\right]$$

Testing the Equality of Population Proportions

$$[\hat{p} \pm Z_{0.025}\sqrt{\hat{p}(1-\hat{p})/n}$$

1.2 Test for Whether A Series Is Normally Distributed

Test for Normal Distribution (Jarque-Bera Test)

$$SK = \frac{E(X-\mu)^2}{\sigma^3}$$

$$K = \frac{E(X-\mu)^4}{\sigma^4}$$

$$MST = \frac{SST}{k-1}$$

$$MSE = \frac{SSE}{n-k}$$

1.3 Parametric Test for Comparing Two Or More Populations

Analysis of Variance (ANOVA)

1.4 Nonparametric Tests for Comparing Two Populations

Testing Location with Independent Samples (Wilcoxon Rank Sum Test)

$$Z = \frac{T_1 - E(T_1)}{\sigma_{T_1}} \sim N(0,1)$$

Testing Location with Dependent Variables – Matched Pairs (Sign Test)

1.5 Nonparametric Test for Comparing Two Or More Populations

Kruskal-Wallis Test

1.6 Correlation

Correlation for Quantitative Data (Pearson Correlation Coefficient)
Correlation for Ordinal Data (Spearman Rank Correlation Coefficient)

2.0 Regression

2.1 Simple Regression

Model

Least squares estimators

Time-Series Data

Cross-Section Data

Constant Error Variance/Homoscedasticity

$$var(\varepsilon_i) = \sigma^2$$

Hypothesis Result

Maximum Sampling Error with Probability $(1-\alpha)$

$$t_{\frac{\alpha}{2},n-2} \times se(\hat{\beta}_j)$$

$100(1-\alpha)\%$ interval estimate/confidence interval for β_j

$$\left[\hat{\beta}_j \pm t_{\frac{\alpha}{2},n-2} \times se(\hat{\beta}_j)\right]$$

Log Linear Equation

Log-Log Equation

2.2 Multiple Regression

Model

Checking for Heteroskedasticity

$$e_i = Y_i - \hat{Y}_i = Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \cdots + \hat{\beta}_{ki} X_k)$$

Goodness of Fit

Testing the Overall Model

Quadratic Models

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_1^2 + \beta_3 X_2 + \beta_4 X_2^2 + \beta_5 X_1 X_2 + \varepsilon$$

$$\frac{dY}{dX_1} = \beta_1 + 2\beta_2 X_1 + \beta_5 X_2 + \varepsilon$$

Hypothesis Testing & Interval Estimation for functions with more than one coefficient

$$var\left(g(\hat{\beta}_2, \hat{\beta}_3)\right) = \left(\frac{dg}{d\hat{\beta}_2}\right)^2 var(\hat{\beta}_2) + \left(\frac{dg}{d\,\hat{\beta}_3}\right)^2 var(\hat{\beta}_3) + 2\left(\frac{dg}{d\hat{\beta}_2}\right)\left(\frac{dg}{d\hat{\beta}_3}\right) cov(\hat{\beta}_2, \hat{\beta}_3)$$

$$\left[\left(-\frac{\hat{\beta}_2}{2\hat{\beta}_3}\right) \pm t_{(n-k-1)} se\left(-\frac{\hat{\beta}_2}{2\hat{\beta}_3}\right)\right]$$

Dummy Variables

Testing Joint Null Hypothesis (Wald F-Test)

Prediction/Forecasting

$$var(Y_0 - \hat{Y}_0) = var(\hat{\beta}_0) + X_0^2 var(\hat{\beta}_1) + 2X_0 cov(\hat{\beta}_0, \hat{\beta}_1) + var(\varepsilon_0)$$

$$var(Y_0 - \hat{Y}_0) = var(\varepsilon_0) = \sigma^2$$

$$se(Y_0 - \hat{Y}_0) = \sqrt{\widehat{var}(Y_0 - \hat{Y}_0)}$$

$$[\hat{Y}_0 \pm t_{(0.025, n-k-1)} se(Y_0 - \hat{Y}_0),]$$

Model Choice Issues

3.0 Binary Choice Models

Linear Probability Model

Probit Model

$$\beta_j \frac{1}{n} \sum \phi(I_i)$$

Logit Model

$$\beta_j \frac{1}{n} \sum \left(\frac{\exp\{-I_i\}}{(1+\exp\{I_i\})^2} \right)$$

4.0 Time Series Regression

Features of Time Series Regressions

Autocorrelations

$$\rho_s = \frac{cov(Y_t, Y_{t-s})}{var(Y_t)}$$

Significance of an Autocorrelation

Correlograms

Time-Series Regression Models

Forecasting From an AR(1) Model

$$\hat{Y}_{T+1} = \hat{\delta} + \hat{\theta}_1 Y_T \qquad \hat{Y}_{T+j} = \hat{\delta} + \hat{\theta}\hat{Y}_{T+j-1}$$

$$\left[\hat{Y}_{T+j} \pm t_{0.025,T-2} se(f_j)\right]$$

AR(1) Model with Trend

$$Y - \mu_Y = \theta_1 (Y_{t-1} - \mu_Y) + \varepsilon_t$$

$$Y_t^* = Y_t - \mu_Y$$

Estimation Strategy

$$\Delta Y_t = \delta + \gamma Y_{t-1} + \varepsilon_t$$

$$\Delta Y_t = \delta + \gamma Y_{t-1} - \theta_2 \Delta Y_{t-1} + \varepsilon_t$$

Two Non-Stationary Models

$$Y_t = \delta + Y_{t-1} + \varepsilon_t$$

Multiplier Analysis

$$Y = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \cdots + v_t$$

$$\frac{dY_t}{dX_t} = \beta_0 = \hat{\delta}_t$$

$$\frac{dY_t}{dX_{t-s}} = \frac{dY_{t+s}}{dX_t} = \hat{\theta}_1^s \hat{\delta}_s$$

$$\sum \frac{dY_t}{dX_{t-s}} = \sum \beta_s = \hat{\delta}_0 + \hat{\delta}_1 + \cdots \hat{\delta}_s$$

$$\sum \frac{dY_t}{dX_{t-j}} = \sum \beta_j = \frac{\hat{\delta}_0}{1 - \hat{\theta}_1}$$

1.0 Hypothesis Testing

Explain the general framework for using data to make inferences about population parameters

The Six Steps of Hypothesis Testing

1. Set up **null** and **alternative hypotheses**
2. **Test statistic and sampling distribution**
3. Specify **significance** level
4. Define **decision rule**
5. Collect sample, compute **test statistic or p-vale**
6. Make **decision** and **conclude**

Rejecting or Accepting the Null

Rejecting the null

- there is not enough evidence to accept the null.

Accepting the null

- **never state the null is true**, simply it is **not false**
- there is not enough evidence to reject the null or
- the evidence from the sample is not compatible with the null

Explain the general framework for using data to make inferences about population parameters

Samples and Data

To learn about **parameters** like μ or p, you use a **sample to make sample statistics** $\bar{X}$ or $\hat{p}$

- **probability distributions** show how good the samples are as estimates

Comparing Two Populations – A Summary

Parametric Tests for Quantitative (Continuous) Data

Independent Samples	Matched Pairs
t-test for equality of means with equal population variances $$t=\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{s_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}\sim t_{(n_1+n_2-2)}$$	*t*-test on sample differences $$t=\frac{\bar{X}_D}{s_D/\sqrt{n}}\sim t_{(n-1)}$$ (makes no assumption about population variances)

Independent Samples	
t-test for equality of means with unequal population variances $$t=\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}\sim t_{(v)}$$	
F-test for equality of variances $$F=\frac{s_1^2}{s_2^2}\sim F_{(n_1-1,\,n_2-1)}$$	

Testing Population Proportions – Categorical Data

Independent Samples	
Testing the equality of two population proportions $$Z=\frac{\hat{p}_1-\hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}\sim N(0,1)$$	Testing a non-zero difference between two population proportions $$Z=\frac{(\hat{p}_1-\hat{p}_2)-D}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}}\sim N(0,1)$$

Nonparametric Tests: Ordinal Data or Quantitative Non-normal Data

Independent samples	Matched Pairs
Wilcoxon rank sum test for locations (or medians) T = sum of ranks $$Z=\frac{T-E(T)}{\sigma_T}\sim N(0,1)$$ Z for large samples;	Sign test for locations (ordinal data) $$Z=\frac{X-0.5n}{0.5\sqrt{n}}\sim N(0,1)$$

1.1 Parametric Tests for Comparing Two Populations

Apply the six hypothesis testing steps to testing a population mean μ

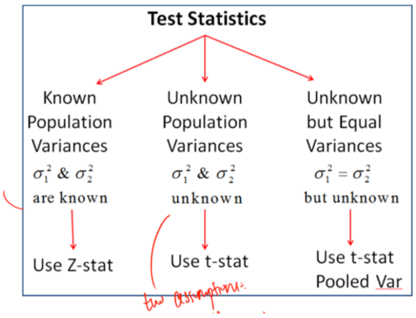
Test hypotheses for the difference between two means using samples of independent quantitative data

Testing Equality of Population Means

(T Test)

$H_0: \mu_1 = \mu_2$ $\qquad$ $H_A: \mu_1 \neq \mu_2 \; or \; \mu_1 < \mu_2 \; or \; \mu_1 > \mu_2$

- when $H_A: \mu_1 \neq \mu_2$, the test is **two-tailed**
  - **p-values on EViews** are always for 2-tail test, must be **halved for 1-tail**
- reject null when t > critical value
- reject null when p < 1 – significance



- **population variances are always known**

95% interval estimate

$$\left[\bar{X} \pm t_{0.025,n-1} \frac{s}{\sqrt{n}}\right]$$

Two t tests for using independent samples of quantitative (continuous) data to test the equality of population means. One test assumes equal variances; the other does not

Independent Samples with Unequal Population Variances

test statistic

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim t(v)$$

degrees of freedom

$$df = \frac{\left[\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right]^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}}$$

Independent Samples and Equal Population Variances

test statistic

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim t_{n_1+n_2-2}$$

pooled variance estimate

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

Estimates common population variance

- $\sigma^2 = \sigma_1^2 = \sigma_2^2$

degrees of freedom

$$v = n_1 + n_2 - 2$$

Explain the difference between testing for the equality of means with independent samples and testing for the equality of means with matched pairs

Test the equality of two means using a sample of matched pairs data

Dependent Samples – Matched Pairs

$H_0: \mu_D = 0$ $H_A: \mu_D \neq 0 \; or \; \mu_D < 0 \; \mu_D > 0$

- matched pairs is better than independent samples as the variance is smaller

test statistic

$$t = \frac{\bar{X}_D}{s_D \sqrt{n}} \sim t_{\alpha, n-1} \; where \; D = X_1 - X_2$$

p-value

$$pr\left(t_n > \frac{\bar{X}_D}{\frac{s_D}{\sqrt{20}}}\right)$$

Apply the six hypothesis testing steps to testing a population variance σ

Test the equality of two population variances

Testing Equality of Population Variances (F Distribution or F Test)

$H_0: \sigma_1^2 = \sigma_2^2$ $H_1: \sigma_1^2 \neq \sigma_2^2 \; or \; \sigma_1^2 < \sigma_2^2 \; or \; \sigma_1^2 > \sigma_2^2$

Tests whether **<u>population variances</u>** are **equal or vary**

- samples are **independent**

Explain how the F-distribution is obtained from two independent chi-square distributions

test statistic or f statstic

$$F = \frac{s_1^2/\sigma_1^2}{s_1^2/\sigma_1^2} \sim F_{n_1 - 1, n_2 - 2}$$

The ratio of two independent chi-square random variables, each divided by its degrees of freedom, is an F random variable or F distribution

- degrees of freedom parameters are $n_2 - 1$ and $n_1 - 1$
- reject H_0 if $F_{1-\frac{\alpha}{2}, n_1 - 1, n_2 - 2} > \frac{s_1^2/\sigma_1^2}{s_1^2/\sigma_1^2} > F_{\frac{\alpha}{2}, n_1 - 1, n_2 - 2}$

confidence interval

$$\left[\frac{(n-1)s^2}{\chi^2_{\alpha\backslash 2,,n-1}}, \frac{(n-1)s^2}{\chi^2_{1-\alpha,n-1}}\right]$$

critical values

$$F_{1-\frac{\alpha}{2},v_1,v_2} = \frac{1}{F_{\frac{\alpha}{2},v_2,v_1}}$$

- lower critical values can be obtained from upper critical values
- alternatively, the larger sample variance can be placed in the numerator.

Apply the six hypothesis testing steps to testing the equality of two population proportions, $H: p_1 = p_2$

Testing the Equality of Population Proportions

$H_0: p_1 = p_2$ $\quad\quad$ $H_1: p_1 \neq p_2 \ or \ p_1 < p_2 \ or \ p_1 > p_2$

Variables are intendent and binomially distributed

- categorical
  - either 1 or 0
- the Satterthwait-Welche t-test on EViews

test statistic

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \sim N(0,1)$$

pooled estimate

$$\hat{p} = \frac{n_1\hat{p} + n_2\hat{p}}{n_1 + n_2}$$

95% interval estimate

$$[\hat{p} \pm Z_{0.025}\sqrt{\hat{p}(1-\hat{p})/n}$$
