Hypothesis Testing I

Sean Hellingman ©

Introduction to Statistical Data Analysis (ADSC1000)

shellingman@tru.ca

Fall 2024

THOMPSON RIVERS UNIVERSITY

Topics

Introduction

- We are continuing to make statistical inferences about the target population.
- Now we are going to draw conclusions about population parameters.
- Example: Made improvements to a pencil making machine. Are pencils actually being produced more quickly?
    - Need statistical methods to draw conclusions about the pace of pencil production.

Hypothesis Testing

- **Hypothesis testing** involves drawing inferences about two contrasting hypotheses (propositions).
  - Related to the value of a population parameter such as mean, proportion, or standard deviation.
- The **null hypothesis** describes an existing theory or a belief.
  - *It takes my pencil making machine at least 4.6 minutes to make a pencil.*
- The **alternative hypothesis** is based on new information provided by sample data.
  - *My improvements mean that it takes **less** than 4.6 minutes to make a pencil.*
- Analogous to: *innocent until proven guilty.*

Steps of Hypothesis Testing

1. Formulate the hypothesis test.
2. Select a *level of significance* (α).
   - **Defines the risk of drawing an incorrect conclusion about the assumed hypothesis that is actually true.**
3. Determining a decision rule on which to base a conclusion.
4. Collecting data and calculating a test statistic.
5. Applying the decision rule to the test statistic in order to draw a conclusion about your hypotheses.

Hypothesis Formulation

- Define two alternative, **mutually exclusive** propositions about one or more population parameters.
- H_0 (**null hypothesis**) represents an existing theory or belief that is accepted to be correct in the absence of contradictory data.
- H_1 (**alternative hypothesis**) is accepted to be true if we reject the null hypothesis.
- Pencil making machine:

$$H_0\text{: } \textit{time to build a pencil} \geq 4.6 \textit{ minutes}$$
$$H_1\text{: } \textit{time to build a pencil} < 4.6 \textit{ minutes}$$

One-Sample Hypothesis Formulations

- **One-Sample Tests**:

H_0: *Population parameter* $\geq$ *constant value*
H_1: *Population parameter* $<$ *constant value*

H_0: *Population parameter* $\leq$ *constant value*
H_1: *Population parameter* $>$ *constant value*

H_0: *Population parameter* $=$ *constant value*
H_1: *Population parameter* $\neq$ *constant value*

Example 1

- Identify the null (H_0) and alternative (H_1) hypotheses for the following situations:
  1. I believe that the modifications to my pencil making machine have actually made production faster than 4.6 minutes per pencil.
  2. I believe that cars actually travel faster than the 50Km/h posted speed limit on McGill road.
  3. An employment company stated that the average income in Kamloops per person is $59 875 per year. I want to check and see if this number is correct.

Two-Sample Hypothesis Formulations

- **Two-Sample Tests**:

H_0: *Population parameter (group 1) - Population parameter (group 2)* ≥ 0

H_1: *Population parameter (group 1) - Population parameter (group 2)* < 0

H_0: *Population parameter (group 1) - Population parameter (group 2)* ≤ 0

H_1: *Population parameter (group 1) - Population parameter (group 2)* > 0

H_0: *Population parameter (group 1) - Population parameter (group 2)* $= 0$

H_1: *Population parameter (group 1) - Population parameter (group 2)* $\neq 0$

Example 2

- Identify the null (H_0) and alternative (H_1) hypotheses for the following situations:
  1. I believe that my new pencil producing machine is faster than my old pencil producing machine.
  2. I believe that cars travel faster on Columbia street than they do on McGill road.
  3. I believe that the average income in Kamloops per person is not the same as the average income in Kelowna per person.

Test Outcomes

- Four possible outcomes of a hypothesis test:
  1. The null hypothesis is true, and the hypothesis test correctly fails to reject it.
  2. The null hypothesis is actually false, and the test correctly rejects the null hypothesis.
  3. The null hypothesis is actually true, but the hypothesis test incorrectly rejects the null hypothesis (**Type I error**).
  4. The null hypothesis is actually false, but the hypothesis test incorrectly fails to reject the null hypothesis (**Type II error**).

Type I Error

- The probability of a Type I error: $P(\textit{Rejecting } H_0 | H_0 \textit{ is True}) = \alpha$
  - Called the **level of significance** of the test.
  - This is the risk you take in making the incorrect conclusion that the alternative hypothesis is true when in fact the null hypothesis is true.

- The **confidence coefficient** is $1 - \alpha$.
  - Probability of correctly failing to reject the null hypothesis.

- Commonly used levels for α: 0.10, 0.05, 0.01
- Confidence levels: 0.90, 0.95, 0.99

Type II Error

- Probability of a Type II error: $P(\textit{Not Rejecting } H_0 | H_0 \textit{ is False}) = \beta$
- We cannot specify β in advance because it depends on an unknown population parameter.
- Generally as α decreases, β increases.
- $1 - \beta$ is called the **power of the test** which is the probability of correctly rejecting the null hypothesis when it is indeed false.
- The power of a test is sensitive to sample size and larger samples can increase the power of tests.

Decision Rules I

- The decision to reject or fail to reject a null hypothesis comes from test statistic calculated using the sample data.

- Is a function of the population parameter of interest and comparing it to a critical value from the hypothesized sampling distribution of the test statistic.

- Sampling distribution is usually the *normal distribution*, *t-distribution*, or some other well known distribution.

- The distribution is divided into two parts: the **rejection region** and the **non-rejection region**.

- If the null hypothesis is false, it is more likely that the test statistic will fall into the rejection region.
    - If the test statistic falls into the rejection region we reject the null hypothesis, otherwise we fail to reject the null hypothesis.

Decision Rules Graphics

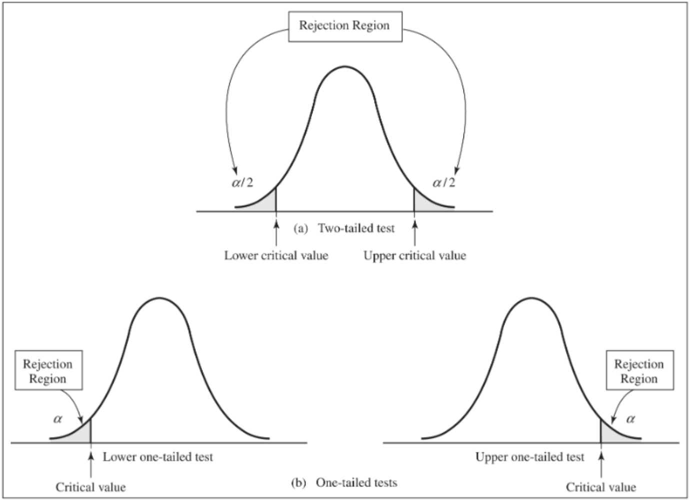


Figure: Source: (1)

Decision Rules II

- The rejection region is defined by the *critical value(s)*.
- Two-tailed tests have two critical values while one-tailed tests only have one critical value.
- Critical values make it *easy* to determine if the test statistic falls within the rejection region or not.
- Examples:
  - For an upper one-tailed test, if the test statistic is greater than the critical value, the decision would be to reject the null hypothesis.
  - For a two-tailed test, if the test statistic is either greater than the upper critical value or less than the lower critical value, the decision would be to reject the null hypothesis.

p-values

- Another way to compare our test statistic to a critical value is by using a *p*-value.

- ***p*-value**: The probability of obtaining a test statistic value equal to or more extreme than that obtained from the sample data when the null hypothesis is true.

- You can think of a *p*-value as an *observed significance level*.

- In other words, a small *p*-value suggests that there is a low probability that the observed differences are due to chance.

- Whenever *p*-value $< \alpha$ reject the null hypothesis.

Example 3

- Do we *reject* or *fail to reject* the null hypothesis with a significance of α=0.05 given the following *p*-values:
  - *p*-value $= 0.560$
  - *p*-value $= 0.042$
  - *p*-value $= 0.065$
  - *p*-value $= 0.002$

One-Sample Hypothesis Tests

Topics

1. Hypothesis tests for means.
2. Hypothesis tests for proportions.
3. Hypothesis tests for variance.

- *All involving only one sample.*

Review

- Recall from confidence intervals for a mean value:
    - If the population standard deviation is known we can use the normal distribution.
    - In practice, the population standard deviation is not known and we use the t-distribution.
- The t-distribution *approaches* the normal distribution as the degrees of freedom (n) increases.

One-Sample t-Tests

- Test statistic:

$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} \tag{1}$$

- Rejection regions:
  - Lower one-tailed test: $t < t_{-\alpha, n-1}$
  - Upper one-tailed test: $t > t_{\alpha, n-1}$
  - Two-tailed test: $|t| > |t_{\alpha/2, n-1}|$

Rejection Regions in R

- $t_{-\alpha,n-1}$ (Lower one-tailed)
    - `qt(p=`α`, df=n-1, lower.tail=TRUE)`
- $t_{\alpha,n-1}$ (Upper one-tailed)
    - `qt(p=`α`, df=n-1, lower.tail=FALSE)`
- $t_{\alpha/2,n-1}$ (Two-tailed)
    - `qt(p=`α`/2, df=n-1, lower.tail=FALSE)`

Example 4

- Do we *reject* or *fail to reject* the null hypothesis with a significance of α=0.05 given the following results (use R):

  - H_0: $\mu \geq 2$ VS H_1: $\mu < 2$
  - Given: $\bar{x} = 1.93$, $s = 1.3$, and $n = 43$

  - H_0: $\mu \leq 1.5$ VS H_1: $\mu > 1.5$
  - Given: $\bar{x} = 1.43$, $s = 0.9$, and $n = 53$

  - H_0: $\mu = 2$ VS H_1: $\neq 2$
  - Given: $\bar{x} = 1.93$, $s = 1.3$, and $n = 43$

p-values in R

- We may also use the `t.test()` function in R.
- This function takes at least one **vector** of values as the first argument.
- Produces the test statistic (t), confidence intervals, *p*-value, and sample mean.
- Usage:
  - Lower one-tailed: `t.test(x, mu =` μ_0`, alternative = "less", conf.level = 0.95)`
  - Upper one-tailed: `t.test(x, mu =` μ_0`, alternative = "greater", conf.level = 0.95)`
  - Two-tailed: `t.test(x, mu =` μ_0`, alternative = "two.sided", conf.level = 0.95)`

Example 5

- Import the *Cars93* dataset from the *MASS* R package into your environment and use the `t.test()` to test the following one-sample hypotheses ($\alpha = 0.05$):
  - The hypothesis that the mean mid-range price (*Price*) of the cars is less than 20 ($20 000).
  - The hypothesis that the mean highway miles per gallon (*MPG.highway*) is greater than 29.
  - That the mean engine size in litres (*EngineSize*) is not 2.4 litres.
- *Hint: Start by identifying* H_0 *and* H_1

Review

- Recall from confidence intervals for a proportion value:
  - An unbiased estimator of a population proportion π is the statistic $\hat{p} = x/n$ (sample proportion).
  - We assume that a sample proportion follows a normal distribution.
- Therefore, we use the quantiles of the standard normal distribution for hypothesis testing of proportions.

One-Sample Tests for Proportions

- Test statistic:

$$z = \frac{\hat{p} - \pi_0}{\sqrt{\pi_0(1 - \pi_0)/n}} \tag{2}$$

- Rejection regions:
  - Lower one-tailed test: $z < z_{-\alpha}$
  - Upper one-tailed test: $z > z_{\alpha}$
  - Two-tailed test: $|z| > |z_{\alpha/2}|$

Rejection Regions in R

- $z_{-\alpha}$ (Lower one-tailed)
  - `qnorm(p=`α`, lower.tail=TRUE)`
- z_{α} (Upper one-tailed)
  - `qnorm(p=`α`, lower.tail=FALSE)`
- $|z_{\alpha/2}|$ (Two-tailed)
  - `qnorm(p=`α`/2, lower.tail=FALSE)`

Example 6

- Do we *reject* or *fail to reject* the null hypothesis with a significance of α=0.05 given the following results (use R):
  - H_0: $\pi \geq 0.65$ VS H_1: $\pi < 0.65$
  - Given: $\hat{p} = 0.63$ and $n = 43$

  - H_0: $\pi \leq 0.5$ VS H_1: $\pi > 0.5$
  - Given: $\hat{p} = 0.43$ and $n = 53$

  - H_0: $\pi = 0.4$ VS H_1: $\neq 0.4$
  - Given: $\hat{p} = 0.44$, and $n = 43$

p-values in R

- We may also use the `prop.test()` function in R.

- This function takes the number of *successes* `x` as the first argument.
- It takes the number of *trials* `n` as the second argument.

- Produces the test statistic, confidence intervals, *p*-value, and sample proportion.

- Usage:
  - Lower one-tailed: `prop.test(x, n, p = ` π_0`, alternative = "less", conf.level = 0.95, correct = FALSE)`
  - Upper one-tailed: `prop.test(x, n, p = ` π_0`, alternative = "greater", conf.level = 0.95, correct = FALSE)`
  - Two-tailed: `prop.test(x, n, p = ` π_0`, alternative = "two.sided", conf.level = 0.95, correct = FALSE)`

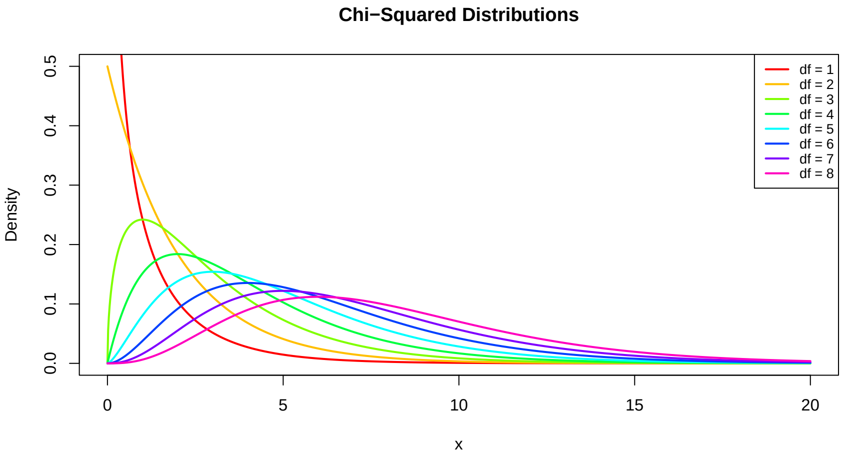
Example 7

- Import the *Cars93* dataset from the *MASS* R package into your environment and use the `prop.test()` to test the following one-sample hypotheses ($\alpha = 0.05$):
    - The hypothesis that the proportion of Small cars (from *Type*) is less than 0.25.
    - The hypothesis that the proportion of Sporty cars (from *Type*) is greater than 0.15.
    - That the proportion of cars on the road with 4 cylinders (*Cylinders*) is not 0.5.
- *Hint: Start by identifying H_0 and H_1*

Variance Review

- Understanding variability is very important in the implementation of statistical theory to decision-making process.
- We have point estimates for variability (standard deviation & variance).
- The sampling distribution of s is **not** normally distributed.
- Instead, the chi-square (χ^2) distribution is used.
- The chi-square (χ^2) distribution also relies on degrees of freedom (df) and it is **not** symmetric.

Chi-square Plots



One-Sample Tests for Variance

- *This test is sensitive to any departures from the normal distribution.*
- Test statistic:

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} \tag{3}$$

- Rejection regions:
  - Lower one-tailed test: $\chi^2 < \chi^2_{1-\alpha,n-1}$
  - Upper one-tailed test: $\chi^2 > \chi^2_{\alpha,n-1}$
  - Two-tailed test: $\chi^2 < \chi^2_{1-\alpha/2,n-1}$ **OR** $\chi^2 > \chi^2_{\alpha/2,n-1}$

Rejection Regions in R

- $\chi^2_{1-\alpha,n-1}$ (Lower one-tailed)
  - qchisq(p=α, df = $n-1$, lower.tail = TRUE)
- $\chi^2_{\alpha,n-1}$ (Upper one-tailed)
  - qchisq(p=α, df = $n-1$, lower.tail = FALSE)
- $\chi^2 < \chi^2_{1-\alpha/2,n-1}$ **OR** $\chi^2 > \chi^2_{\alpha/2,n-1}$ (Two-tailed)
  - qchisq(p=$\alpha/2$, df = $n-1$, lower.tail = TRUE) **AND**
  - qchisq(p=$\alpha/2$, df = $n-1$, lower.tail = FALSE)

p-values in R

- We may also use the `varTest()` function in the *EnvStats* R package.
- This function takes a numeric vector `x` as the first argument.
- Produces the test statistic, confidence intervals, *p*-value, and sample variance.
- Usage:
  - Lower one-tailed: `varTest(x, alternative = "less", sigma.squared =` σ_0^2`, conf.level = 0.95)`
  - Upper one-tailed: `varTest(x, alternative = "greater", sigma.squared =` σ_0^2`, conf.level = 0.95)`
  - Two-tailed: `varTest(x, alternative = "two.sided", sigma.squared =` σ_0^2`, conf.level = 0.95)`

Example 8

- Import the *swiss* dataset from base R into your environment and use the `varTest()` from the *EnvStats* R package to test the following one-sample hypotheses ($\alpha = 0.05$):
  - The hypothesis that the variance of the draftees receiving highest mark on army examination (*Examination*) is less than 66.
  - The hypothesis that the variance of the fertility (*Fertility*) is greater than 145.
  - That the variance of (*Examination*) is not 70.
- *Hint: Start by identifying H_0 and H_1*

Power of a Test

- $1 - \beta$ is called the **power of the test** which is the probability of correctly rejecting the null hypothesis when it is indeed false.
- The power of a test depends on the true value of the population mean, the level of confidence used, and the sample size.
- Assuming that H_1 is true, there will be some overlap in the sampling distributions of μ_0 and μ_1.
- This means that the test statistic may fall into the *acceptance* region even when H_1 is true.

Power of a Test Graphics I

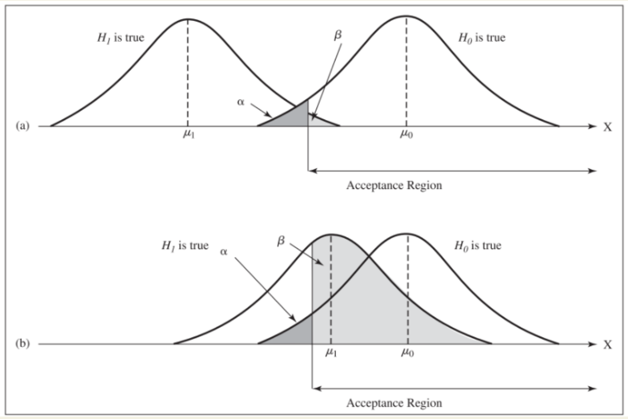


Figure: Source: (1)

Power of a Test Graphics II

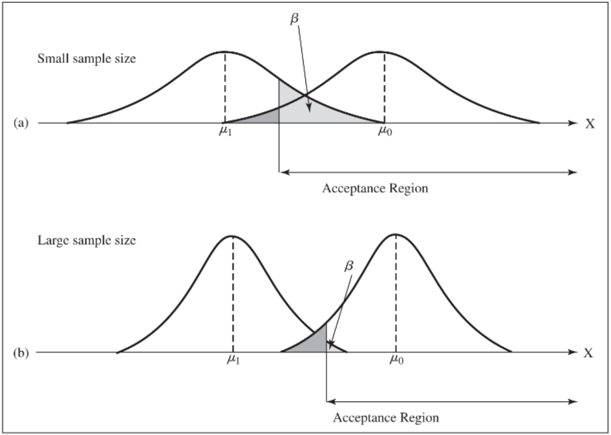


Figure: Source: (1)

Exercise 1

- Import the *Cars93* dataset from the *MASS* R package into your environment and test the following one-sample hypotheses ($\alpha = 0.01$):
  - That the mean passenger capacity (*Passengers*) is not 4.
  - The hypothesis that the mean engine revolutions per mile (*Rev.per.mile*) is greater than 2100.
  - The hypothesis that the mean horsepower (*Horsepower*) of the cars is less than 155.
- *Hint: Start by identifying H_0 and H_1*

Exercise 2

- Import the *swiss* dataset into your environment and test the following one-sample hypotheses ($\alpha = 0.05$):
  - The hypothesis that the proportion of Large cars (from *Type*) is less than 0.20.
  - The hypothesis that the proportion of non-USA or USA company origins (*Origin*) is not the same.
  - That the proportion of cars on the road with 6 cylinders (*Cylinders*) is greater than 0.25.
- *Hint: Start by identifying H_0 and H_1*

Exercise 3

- Import the *Cars93* dataset from the *MASS* R package into your environment and test the following one-sample hypotheses ($\alpha = 0.10$):
    - The hypothesis that the variance of the draftees receiving highest mark on army examination (*Examination*) is greater than than 60.
    - The hypothesis that the variance of the fertility (*Fertility*) is not 145.
    - That the variance of (*Examination*) is less than 70.
- *Hint: Start by identifying H_0 and H_1*

References & Resources

1. Evans, J. R., Olson, D. L., & Olson, D. L. (2007). *Statistics, data analysis, and decision modeling.* Upper Saddle River, NJ: Pearson/Prentice Hall.

2. Devore, J. L., Berk, K. N., & Carlton, M. A. (2012). *Modern mathematical statistics with applications (Second Edition).* New York: Springer.

- One-Sample Chi-Squared Test on Variance
- Student's t-Test
- Test of Equal or Given Proportions