Hypothesis Testing (Ch. 6.2)

Will Landau

Iowa State University

Apr 2, 2013

Outline

Statistical inference

- **Statistical inference**: using data from the sample to draw conclusions about the population
  - Point estimation (confidence intervals): estimating population parameters and specifying the degree of precision of the estimate.
  - Hypothesis testing: testing the validity of statements about the population that are framed in terms of parameters.

Hypothesis testing

- **Hypothesis testing (significance testing)**: the use of data in the quantitative assessment of the plausibility of some trial value or a parameter.
- You have competing **hypotheses**, or statements, about a population:
  - The **null hypothesis**, denoted H_0 is the proposition that a parameter equals some fixed number.
  - The **alternative hypothesis**, denoted H_a or H_1, is a statement that stands in opposition to the null hypothesis.
  - Examples:

$$\begin{array}{lll} \mathrm{H_0}: \mu = \# & \mathrm{H_0}: \mu = \# & \mathrm{H_0}: \mu = \# \\ \mathrm{H_a}: \mu > \# & \mathrm{H_a}: \mu < \# & \mathrm{H_a}: \mu \neq \# \end{array}$$

- Note: $H_a : \mu \neq \#$ makes a **two-sided test**, while $H_a : \mu < \#$ and $H_a : \mu > \#$ make a **one-sided test**.
- The goal is to use the data to debunk the null hypothesis in favor of the alternative:
  - Assume H_0.
  - Try to show that, under H_0, the data are preposterous.
  - If the data are preposterous, reject H_0 and conclude H_a.

Hypothesis testing

- Outcomes of a hypothesis test:

The ultimate decision is in favor of:

The true state of affairs is described by:	H_0	H_a
H_0		Type I error
H_a	Type II error	

- α (the very same α in confidence intervals) is the probability of rejecting H_0 when H_0 is true.
  - α is the Type I Error probability.
  - For honesty's sake, α is fixed before you even *look* at the data.

Formal steps of a hypothesis test using confidence intervals

1. State H_0 and H_a.
2. State α.
3. State the form of the $1 - \alpha$ confidence interval you will use, along with all the assumptions necessary.
4. Calculate the $1 - \alpha$ confidence interval.
5. Based on the $1 - \alpha$ confidence interval, either:
   - Reject H_0 and conclude H_a, or
   - Fail to reject H_0.
6. Interpret the conclusion using layman's terms.

Example: breaking strength of wire

- Suppose you are a manufacturer of construction equipment. You make 0.0125 inch wire rope and need to determine how much weight it can hold before breaking so that you can label it clearly.
- Here are breaking strengths, in kg, for 40 sample wires:

100.37	96.31	72.57	88.02	105.89	107.80	75.84	92.73	67.47
94.87	122.04	115.12	95.24	119.75	114.83	101.79	80.90	96.10
118.51	109.66	88.07	56.29	86.50	57.62	74.70	92.53	86.25
82.56	97.96	94.92	62.93	98.44	119.37	103.70	72.40	71.29
107.24	64.82	93.51	86.97					

- Let's conduct a hypothesis test to find out if the true mean breaking strength is above 85 kg.

Example: breaking strength of wire

1. $H_0 : \mu = 85$ kg and $H_a : \mu > 85$ kg, where μ is the true mean breaking strength.
2. $\alpha = 0.05$
3. Since this is a one-sided (lower) test, I will use a lower $1 - \alpha$ confidence interval:

$$\left(\overline{x} - z_{1-\alpha} \frac{s}{\sqrt{n}}, \ \infty \right)$$

   I am assuming:

   - The data points $x_1, \dots x_n$ were iid draws from some distribution with mean μ and some constant variance.

4. From before, we calculated the confidence interval to be $(87.24, \infty)$.
5. With 95% confidence, we have shown that $\mu > 87.24$. Hence, at significance level $\alpha = 0.05$, we have shown that $\mu > 85$. We reject H_0 and conclude H_a.
6. There is enough evidence to conclude that the true mean breaking strength of the wire is greater than 85 kg. Hence, the requirement is met.

Outline

Hypothesis testing with critical values

- Instead of using a confidence interval in the test, simply compute a test statistic and compare it to a **critical value**.
- A **test statistic** is a number of the form:

$$K = \frac{\overline{x} - \mu_0}{\phi}$$

  - μ_0 is the true mean value of the data under the null hypothesis.
  - ϕ is either $\sigma/\sqrt{n}$ or $s/\sqrt{n}$, whichever version of $SD(\overline{X})$ is available.
- A **critical value** is a special quantile on the distribution of K (either $z_{1-\alpha}$, $z_{1-\alpha/2}$, $t_{n-1,1-\alpha}$, or $t_{n-1,1-\alpha/2}$). We compare it to K to decide whether to reject H_0 or fail to reject H_0.

Full list of steps: critical values

1. State H_0 and H_a.
2. State α.
3. State the form of the test statistic, its distribution under the null hypothesis, and all your assumptions.
4. Calculate the test statistic and the critical value
5. Based on the previous step, either:
   - Reject H_0 and conclude H_a, or
   - Fail to reject H_0.
6. Interpret the conclusion using layman's terms.

Example: fill weight of jars

- Suppose a manufacturer fills jars of food using a stable filling process with a known standard deviation of $\sigma = 1.6g$.
- We take a sample of $n =$47 jars and measure the sample mean weight $\overline{x} = 138.2$ g.
- I will conduct the following hypothesis tests:
  - $H_0 : \mu = 140$ vs. $H_a : \mu \neq 140$
  - $H_0 : \mu = 138$ vs. $H_a : \mu < 138$

$H_0 : \mu = 140$ vs. $H_a : \mu \neq 140$

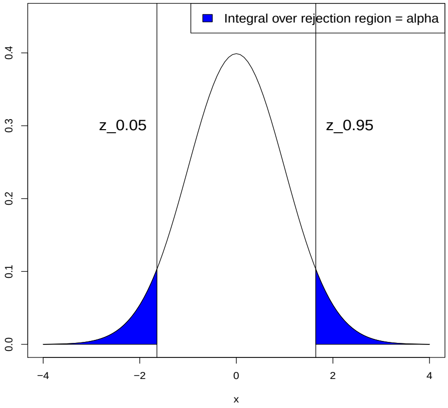
1. $H_0 : \mu = 140$, $H_a : \mu \neq 140$
2. $\alpha = 0.1$
3. Since σ is known and the sample size is large enough for the Central Limit Theorem, I will use the test statistic:

$$K = \frac{\overline{x} - 140}{\sigma/\sqrt{n}}$$

- Assume $X_1, \ldots, X_n$ are iid with mean μ and variance σ^2.
- $K \sim N(0, 1)$ under the null hypothesis.
- Since $K \sim N(0, 1)$ and this is a 2-sided test, I reject H_0 when $|K| > |z_{1-\alpha/2}|$.

$H_0 : \mu = 140$ vs. $H_a : \mu \neq 140$

- **Rejection region**: the set of all possible values of K for which the H_0 is rejected.
- The pdf of K must integrate to α over the rejection region (in this case, $(-\infty, z_{\alpha/2})$ and $(z_{1-\alpha/2},\ \infty)$).



$H_0 : \mu = 140$ vs. $H_a : \mu \neq 140$

4. The moment of truth:
   - $$K = \frac{138.2 - 140}{1.6/\sqrt{47}} = -7.72$$
   - $z_{1-\alpha/2} = z_{1-0.1/2} = z_{0.95} = 1.64.$
5. Since $|K| = |-7.72| > 1.64 = |z_{1-\alpha/2}|$, I reject H_0 in favor of H_a.
6. There is strong evidence that the true mean fill weight is not 140 g.

$H_0 : \mu = 138$ vs. $H_a : \mu < 138$

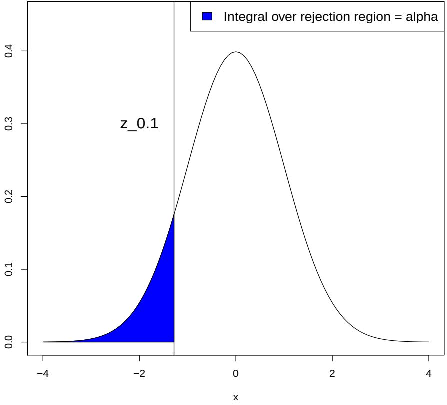
1. $H_0 : \mu = 138$, $H_a : \mu < 138$
2. $\alpha = 0.1$
3. Since σ is known and the sample size is large enough for the Central Limit Theorem, I will use the test statistic:

$$K = \frac{\overline{x} - 138}{\sigma/\sqrt{n}}$$

- Assume $X_1, \ldots, X_n$ are iid with mean μ and variance σ^2.
- $K \sim N(0,1)$ under the null hypothesis.
- Since $K \sim N(0,1)$ and this is a 1-sided upper test, I reject H_0 when $K < z_\alpha$.

$H_0 : \mu = 138$ vs. $H_a : \mu < 138$

- This time, our rejection region is $(-\infty,\ z_\alpha)$.
- The pdf of K must integrate to α over the rejection region.



$H_0 : \mu = 138$ vs. $H_a : \mu < 138$

4. The moment of truth:
   - $$K = \frac{138.2 - 138}{1.6/\sqrt{47}} = 0.857$$
   - $z_\alpha = z_{0.1} = -1.28.$
5. Since $K = 0.857$, which is not less than $z_\alpha = -1.28$, I fail to reject H_0.
6. There is not enough evidence to conclude that the true mean fill weight is less than 138 g.

Example: concrete beams

- 10 concrete beams were each measured for flexural strength (MPa):

8.2	8.7	7.8	9.7	7.4
7.8	7.7	11.6	11.3	11.8

- $\overline{x} = 9.2$ MPa, $s = 1.76$ MPa.
- I will conduct a hypothesis test to find out if the flexural strength is above 8.0 MPa.

Example: concrete beams

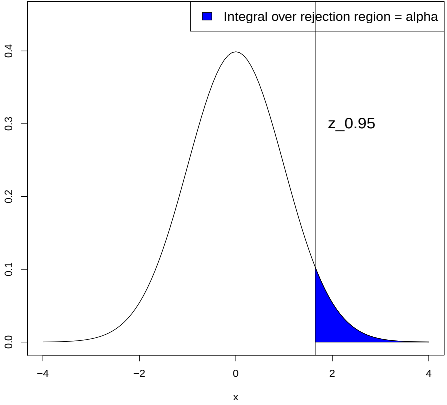
1. $H_0 : \mu = 8.0$, $H_a : \mu > 8.0$
2. $\alpha = 0.05$
3. Since the sample size is small, I will use the test statistic:

$$K = \frac{\overline{x} - 8.0}{s/\sqrt{n}}$$

- Assume $X_1, \dots, X_n$ are iid $N(\mu, \sigma^2)$
- $K \sim t_{n-1} = t_9$ under the null hypothesis because n is small and σ is unknown.
- Since $K \sim t_9$ and this is a 1-sided lower test, I reject H_0 when $K > t_{9,1-\alpha}$.

Example: concrete beams

- This time, our rejection region is $(z_{1-\alpha}, \infty)$.
- The pdf of K must integrate to α over the rejection region.



Example: concrete beams

4. The moment of truth:
   - $$K = \frac{9.2 - 8.0}{1.76/\sqrt{10}} = 2.16$$
   - $z_{1-\alpha} = z_{0.95} = 1.64$.
5. Since $K = 2.16 > z_{\alpha} = 1.64$, I reject H_0 in favor of H_a.
6. There is enough evidence to conclude that the true mean flexural strength of the beams is above 8.0 MPa.

Which test statistics and critical values to use

- The rules for test statistics depend on the sample size n and the knowledge of σ in the same way confidence intervals do.

Condition	Test Statistic K	Distribution of K
$n \geq 25$, σ known	$\frac{\mu - \mu_0}{\sigma/\sqrt{n}}$	$N(0,1)$
$n \geq 25$, σ unknown	$\frac{\mu - \mu_0}{s/\sqrt{n}}$	$N(0,1)$
$n < 25$, σ unknown	$\frac{\mu - \mu_0}{s/\sqrt{n}}$	t_{n-1}

- Appropriate comparisons of critical values with the test statistic:

	$H_a : \mu \neq \mu_0$	$H_a : \mu < \mu_0$	$H_a : \mu > \mu_0$
$n \geq 25, \sigma$	$\lvert K \rvert > \lvert z_{1-\alpha/2} \rvert$	$K < z_\alpha$	$K > z_{1-\alpha}$
$n \geq 25, s$	$\lvert K \rvert > \lvert z_{1-\alpha/2} \rvert$	$K < z_\alpha$	$K > z_{1-\alpha}$
$n < 25, s$	$\lvert K \rvert > \lvert t_{n-1,\ 1-\alpha/2} \rvert$	$K < t_{n-1,\ \alpha}$	$K > t_{n-1,\ 1-\alpha}$

Your turn: car engines

- Consider a grinding process used to rebuild car engines, which involves grinding rod journals for engine crankshafts.
- Of interest is the deviation of the true mean rod journal diameter from the target diameter.
- 32 consecutive rod journals are ground, with a sample mean deviation of -0.16×10^{-4} in from the target diameter.
- The sample standard deviation of these deviations is $s = 0.7 \times 10^{-4}$ in.
- At a significance level of $\alpha = 0.05$, conduct a hypothesis test to determine whether the rod journal diameters are significantly off target.

Answers: car engines

1. $H_0 : \mu = 0$, $H_a : \mu \neq 0$.
2. $\alpha = 0.05$
3. Since σ is unknown, I use:

$$K = \frac{\overline{x} - 8.0}{s/\sqrt{n}}$$

- Assume $X_1, \ldots, X_n$ are iid (μ, σ^2). Since $n \geq 25$, they don't need to be normally distributed.
- $K \sim N(0,1)$ under the null hypothesis because $n \geq 25$.
- Since $K \sim N(0,1)$ and this is a 2-sided test, I reject H_0 when $|K| > |z_{1-\alpha/2}|$.

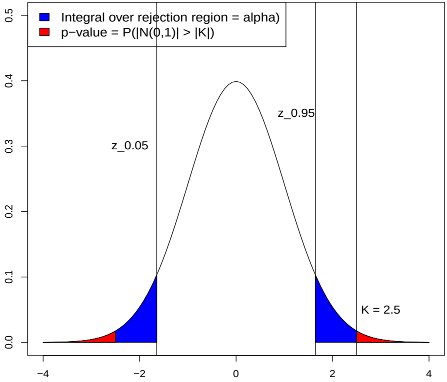
Answers: car engines

4. The moment of truth:
   - $K = \frac{-0.16 \times 10^{-4} - 0}{0.7 \times 10^{-4} / \sqrt{32}} = -1.29$
   - $z_{1-\alpha/2} = z_{0.975} = 1.96$.
5. Since $|K| = 1.29 \not> z_\alpha = 1.96$, I fail to reject H_0.
6. There is not enough evidence to conclude that the rod journal diameters are off target.

Outline

p-values

- A **p-value** is the probability of getting a result at least as extreme as the one observed under the null hypothesis.
- More specifically, it's the probability (assuming the null hypothesis is true) of observing a test statistic farther into the rejection region than K.



Full list of steps: p-values

1. State H_0 and H_a.
2. State α.
3. State the form of the test statistic, its distribution under the null hypothesis, and all your assumptions.
4. Calculate the test statistic and the p-value
5. Make a decision based on the p-value.
   - If the p-value $< \alpha$, reject H_0 and conclude H_a.
   - Otherwise, fail to reject H_0.
6. Interpret the conclusion using layman's terms.

Calculating p-values

- Let K be the value of the test statistic, $Z \sim N(0, 1)$, and $T \sim t_{n-1}$. Here is a table of p-values that you should use for each set of conditions and choice of H_a.

	$H_a : \mu \neq \mu_0$	$H_a : \mu < \mu_0$	$H_a : \mu > \mu_0$
$n \geq 25, \sigma$	$P(\|Z\| > \|K\|)$	$P(Z < K)$	$P(Z > K)$
$n \geq 25, s$	$P(\|Z\| > \|K\|)$	$P(Z < K)$	$P(Z > K)$
$n < 25, s$	$P(\|T\| > \|K\|)$	$P(T < K)$	$P(T > K)$

Example: concrete beams

- 10 concrete beams were each measured for flexural strength (MPa):

8.2	8.7	7.8	9.7	7.4
7.8	7.7	11.6	11.3	11.8

- $\overline{x} = 9.2$ MPa, $s = 1.76$ MPa.
- I will conduct a hypothesis test to find out if the flexural strength is different from 9.0 MPa.

Example: concrete beams

1. $H_0 : \mu = 9.0$, $H_a : \mu \neq 9.0$
2. $\alpha = 0.05$
3. Since the sample size is small, I will use the test statistic:

$$K = \frac{\overline{x} - 9.0}{s/\sqrt{n}}$$

- Assume $X_1, \ldots, X_n$ are iid $N(\mu, \sigma^2)$
- $K \sim t_{n-1} = t_9$ under the null hypothesis because n is small and σ is unknown.

Example: concrete beams

4. The moment of truth:
   - $$K = \frac{9.2 - 9.0}{1.76/\sqrt{10}} = 0.359$$
   - p-value:

$$\begin{aligned} P(|t_9| > 0.359) &= P(t_9 > 0.359) + P(t_9 < -0.359) \\ &= 1 - P(t_9 \leq 0.359) + P(t_9 < -0.359) \\ &= 1 - 0.64 + 0.36 \\ &= 0.72 \end{aligned}$$

5. Since the p-value $= 0.72 > \alpha$, I fail to reject H_0.
6. There is not enough evidence to conclude that the true mean flexural strength of the beams is different from 9.0 MPa.

Your turn: cylinders

- The strengths of 40 steel cylinders were measured in MPa.
- The sample mean strength is 1.2 MPa with a sample standard deviation of 0.5 MPa.
- At significance level $\alpha = 0.01$, conduct a hypothesis test to determine if the cylinders meet the strength requirement of 0.8 MPa.

Answers: cylinders

1. $H_0 : \mu = 1.0$, $H_a : \mu > 0.8$.
2. $\alpha = 0.01$.
3. Since σ is unknown, I use the test statistic:

$$K = \frac{\overline{x} - 0.8}{s/\sqrt{n}}$$

- I assume $X_1, \ldots, X_{40}$ are iid with mean μ and variance σ^2.
- $K \sim N(0,1)$ by the Central Limit Theorem since n is large.

Answers: cylinders

4. The moment of truth:
   - $$K = \frac{1.2 - 0.8}{0.5/\sqrt{40}} = 5.06$$
   - p-value:
   $$\begin{aligned} P(Z > 5.06) &= 1 - P(Z \leq 5.06) \\ &= 1 - \Phi(5.06) \\ &\approx 1 - 1 \\ &= 0 \end{aligned}$$
5. Since the p-value $<< \alpha$, I reject H_0 and conclude H_a.
6. There is overwhelming evidence to conclude that the cylinders meet the strength requirement of 0.8 MPa.