Hypothesis Tests
Part 2: One-Sample Tests about Central Tendency

DASC 512

Overview

- Large-Sample Test about Mean/Proportion (*z* test)
- One-Sample Test about Mean (*t* test)
- One-Sample Test about Median (Wilcoxon Signed Rank Sum)
- One-Sample Test about Proportion (Binomial test)

What are these different tests?

- Fundamentally, hypothesis tests are the same
  - They test whether the evidence supports an alternative hypothesis
- Each test is based on different assumptions
  - Number of samples/treatments
  - Parameter of interest
  - Distributional assumptions
  - Sample size

Overview

- Large-Sample Test about Mean/Proportion (*z* test)
- One-Sample Test about Mean (*t* test)
- One-Sample Test about Median (Wilcoxon Signed Rank Sum)
- One-Sample Test about Proportion (Binomial test)

Assumptions

- Type of data – Numerical
- Randomization – Data gathered randomly (iid)
- Population distribution – Assumes population data distributed $N(\mu, \sigma^2)$, but test is robust to deviations thanks to Central Limit Theorem
  - A method is said to be robust with respect to an assumption if it performs adequately even when that assumption is violated.
- Sample size – Used only for large sample sizes (think CLT)

$$(n \geq 30, np > 10, n(1-p) > 10)$$

Hypotheses

Null hypothesis: The mean is equal to some known value

$$H_0: \mu = \mu_0$$

Alternative hypothesis: The mean is not equal to the known value in some way (depends on the research hypothesis)

$$H_a: \mu < \mu_0, \qquad H_a: \mu > \mu_0, \qquad H_a: \mu \neq \mu_0$$

Test Statistic

Assuming the population is distributed $N(\mu, \sigma^2)$, the sampling distribution is $N\left(\mu, \frac{\sigma^2}{n}\right)$. We'll normalize our results to the sampling distribution.

The test statistic is then

$$Z = \frac{(\bar{X} - \mu_0)}{SEM}$$

where

$$SEM = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}} \approx \frac{s}{\sqrt{n}}$$

Conclusion

Rejection region varies on the alternative hypothesis.

If $H_a\!: \mu < \mu_0$,

$$Z < N_{PPF}(\alpha)$$

If $H_a\!: \mu > \mu_0$,

$$Z > N_{ISF}(\alpha)$$

If $H_a\!: \mu \neq \mu_0$,

$$|Z| > N_{ISF}\left(\frac{\alpha}{2}\right)$$

Conclusion

If the observed value falls within the rejection region,

We reject H_0 at the α level of significance.

If the observed value falls outside the rejection region,

We fail to reject H_0 at the α level of significance.

P-value

Presuming that H_0 is true, the probability of observing a test statistic (Z) more extreme than the one observed (z) is:

If $H_a\!: \mu < \mu_0$,

$$p = P(\mathrm{Z} < \mathrm{z}) = N_{CDF}(z)$$

If $H_a\!: \mu > \mu_0$,

$$p = P(\mathrm{Z} > \mathrm{z}) = N_{SF}(z)$$

If $H_a\!: \mu \neq \mu_0$,

$$p = P(|Z| > |z|) = 2 \times N_{SF}(|z|)$$

Example: General Social Survey (GSS)

Political Ideology	Race		
	Hispanic	Black	White
1. Extremely liberal	5	16	73
2. Liberal	49	52	209
3. Slightly liberal	46	42	190
4. Moderate, middle of the road	155	182	705
5. Slightly conservative	50	43	260
6. Conservative	50	25	314
7. Extremely conservative	14	11	84
n	369	371	1835

https://sda.Berkeley.edu/archive.htm

Example: General Social Survey (GSS)

- Research Question: Are Hispanic Americans, on average, moderate?

- Assumptions:
  - Data is quantitative.
  - Data is gathered randomly.
  - Population data is normally distributed.
  - Sample size is large. $n = 369$

https://sda.Berkeley.edu/archive.htm

Example: General Social Survey (GSS)

Hypotheses:

$$H_0: \mu = 4.0$$
$$H_a: \mu \neq 4.0$$

https://sda.Berkeley.edu/archive.htm

Example: General Social Survey (GSS)

Observed mean value:

$$\bar{x} = \frac{\sum_{i=1}^{7} i \times n_i}{n}$$

where n_i is the number of observations of value i.

https://sda.Berkeley.edu/archive.htm

Example: General Social Survey (GSS)

Test statistic:

$$z = \frac{\bar{x} - 4.0}{\mathrm{SEM}} = \frac{(\bar{x} - 4.0)\sqrt{369}}{s_x}$$

$$s_x = \sqrt{\frac{\sum_{i=1}^{7} n_i (i - \bar{x})^2}{n - 1}}$$

https://sda.Berkeley.edu/archive.htm

Example: General Social Survey (GSS)

Rejection region:

$$|z| > N_{ISF}\left(\frac{\alpha}{2}\right)$$

Let $\alpha = 0.05$ for this test.

$$|z| > N_{ISF}(0.025) = 1.96$$

https://sda.Berkeley.edu/archive.htm

Example: General Social Survey (GSS)

P-value:

$$p = P(|Z| > |z|) = 2 \times N_{ISF}(|z|)$$

Confidence interval:

$$\bar{x} \pm N_{ISF}\left(\frac{\alpha}{2}\right) \times SE$$

https://sda.Berkeley.edu/archive.htm

Example: General Social Survey (GSS)

<u>Conclusion:</u>

We fail to reject the hypothesis at significance $\alpha = 0.05$ that Hispanic Americans have a mean political ideology of 4.0 (moderate, middle of the road).

The observed p-value was $\mathrm{p} = 0.20$.

We can conclude with 95% confidence that the mean political ideology of Hispanic Americans is between 3.95 and 4.23.

https://sda.Berkeley.edu/archive.htm

Z-test for proportions

If the parameter of interest is π, the proportion of events that are successes (a binomial test), the z-test can be used if

$$\max\big(np, n(1-p)\big) \geq 10$$

Hypotheses:

$$H_0: \pi = \pi_0, \qquad H_a: \pi < \pi_0, \qquad \pi > \pi_0, \qquad \pi \neq \pi_0$$

Test statistic:

$$Z = \frac{p - \pi_0}{se_0}, \qquad \text{where } se_0 = \sqrt{\frac{\pi_0(1-\pi_0)}{n}}$$

It is otherwise executed as if for a mean.

When (not) to use a z-test

- Traditionally, z tests allowed usage of probability tables
- With modern software, the z test is nearly obsolete
  - Some textbooks only include it for testing proportions
  - Scipy.stats doesn't even have a z-test function
- For one-sample tests about a mean, use the t-test (next)
  - Scipy.stats: stats.ttest_1samp(a, popmean, axis, nan_policy, alternative)
- For one-sample tests of proportion, use the binomial test (soon)
  - Scipy.stats: stats.binom_test(x, n, p, alternative)
- We'll talk about two-sample tests later… and alternatives to z.

Overview

- Large-Sample Test about Mean/Proportion (z test)
- One-Sample Test about Mean (t test)
- One-Sample Test about Median (Wilcoxon Signed Rank Sum)
- One-Sample Test about Proportion (Binomial test)

Assumptions

- Type of data – Quantitative
- Randomization – Data gathered randomly (iid)
- Population distribution – Assumes population data distributed $N(\mu, \sigma^2)$, but test may be robust depending on actual distribution and sample size
- Sample size – Can be used for small sample size

Hypotheses

Null hypothesis: The mean is equal to some known value

$$H_0: \mu = \mu_0$$

Alternative hypothesis: The mean is not equal to the known value in some way (depends on the research hypothesis)

$$H_a: \mu < \mu_0, \qquad H_a: \mu > \mu_0, \qquad H_a: \mu \neq \mu_0$$

Test Statistic

Assuming the population is distributed $N(\mu, \sigma^2)$, the sampling distribution is

$$\bar{X} \sim \mu + \frac{\sigma}{\sqrt{n}} t^*(\nu = n - 1)$$

We'll normalize our results to the sampling distribution. The test statistic is then

$$T = \frac{(\bar{X} - \mu_0)}{SE}$$

where

$$SE = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}} \approx \frac{s}{\sqrt{n}}$$

Conclusion

- Rejection region varies on the alternative hypothesis.

- If $H_a: \mu < \mu_0$,

$$T < t_{PPF}(\alpha, \nu = n - 1)$$

- If $H_a: \mu > \mu_0$,

$$T > t_{ISF}(\alpha, \nu = n - 1)$$

- If $H_a: \mu \neq \mu_0$,

$$|T| > t_{ISF}\left(\frac{\alpha}{2}, \nu = n - 1\right)$$

Conclusion

- If the observed value falls within the rejection region,

  We reject H_0 at the α level of significance.

- If the observed value falls outside the rejection region,

  We fail to reject H_0 at the α level of significance.

P-value

Presuming that H_0 is true, the probability of observing a test statistic (T) more extreme than the one observed (t) is:

- If $H_a : \mu < \mu_0$,

$$p = P(\mathrm{T} < \mathrm{t}) = t_{CDF}(t|\nu = n - 1)$$

- If $H_a : \mu > \mu_0$,

$$p = P(\mathrm{T} > \mathrm{t}) = t_{ISF}(t|\nu = n - 1)$$

- If $H_a : \mu \neq \mu_0$,

$$p = P(|T| > |t|) = 2 \times t_{ISF}(|t||\nu = n - 1)$$

Overview

- Large-Sample Test about Mean/Proportion (*z* test)
- One-Sample Test about Mean (*t* test)
- One-Sample Test about Median (Wilcoxon Signed Rank Sum)
- One-Sample Test about Proportion (Binomial test)

Assumptions

- Type of data – Quantitative
- Randomization – Data gathered randomly (iid)
- Population distribution – No assumptions
- Sample size – Can be used for small sample size

Hypotheses

Null hypothesis: The data is symmetric about some known median value

$$H_0: \theta = \theta_0$$

Alternative hypothesis: The median is not equal to the known value in some way (depends on the research hypothesis)

$$H_a: \theta < \theta_0, \qquad H_a: \theta > \theta_0, \qquad H_a: \theta \neq \theta_0$$

Test Statistic

For each observation x_i, let $d_i = x_i - \theta_0$. Rank order the data from smallest to largest according to $|d_i|$.

- Let W^+ be the sum of the ranks of observations with $x_i > 0$.
- Let W^- be the sum of the ranks of observations with $x_i < 0$.

Our test statistic is

$$W = \min(W^-, W^+)$$

Test Statistic Example

A rifle marksman takes note of her horizontal miss distance shooting at a target 8 times. Are her shots centered on the target?

Distance (inches)
-1.1
1.9
-0.6
1.6
-1.2
-3.2
-3.4
-1.9

Test Statistic Example

Because the expected median is $\theta = 0$, $x_i = d_i$. We can then calculate the absolute deviation from the expected median.

Distance (inches)	d_i	$\lvert d_i \rvert$	Rank
-1.1	-1.1	1.1	
1.9	1.9	1.9	
-0.6	-0.6	0.6	
1.6	1.6	1.6	
-1.2	-1.2	1.2	
-3.2	-3.2	3.2	
-3.4	-3.4	3.4	
-1.9	-1.9	1.9	

Test Statistic Example

Now we give each observation a rank. In case of ties, we assign the average rank to all tied values.

Distance (inches)	d_i	$\lvert d_i \rvert$	Rank
-1.1	-1.1	1.1	2
1.9	1.9	1.9	5.5
-0.6	-0.6	0.6	1
1.6	1.6	1.6	4
-1.2	-1.2	1.2	3
-3.2	-3.2	3.2	7
-3.4	-3.4	3.4	8
-1.9	-1.9	1.9	5.5

Test Statistic Example

- $W^+ = 5.5 + 4 = 9.5,\ W^- = 2 + 1 + 3 + 7 + 8 + 5.5 = 26.5$
- $W = \min(W^-, W^+) = 9.5$

Distance (inches)	d_i	$\lvert d_i \rvert$	Rank
-1.1	-1.1	1.1	2
1.9	1.9	1.9	5.5
-0.6	-0.6	0.6	1
1.6	1.6	1.6	4
-1.2	-1.2	1.2	3
-3.2	-3.2	3.2	7
-3.4	-3.4	3.4	8
-1.9	-1.9	1.9	5.5

Conclusion

The p-value and conclusion use the Wilcoxon T distribution, which is only used for this test (and its two-sample counterpart) and has no closed form.

For large samples, this distribution is computationally intractable, and the normal approximation (z test) is used. Scipy.stats automatically does this.

Conclusion

So $p = 0.25$. We fail to reject the null hypothesis.

Why Wilcoxon?

First, to introduce you to non-parametric statistics, which don't rely upon distributional assumptions and instead use fields such as order statistics and the binomial distribution.

Second, the Wilcoxon signed-rank sum test is a powerful non-parametric test analogous to the t-test. With small samples and highly skewed data, it is a great choice.

Overview

- Large-Sample Test about Mean/Proportion (*z* test)
- One-Sample Test about Mean (*t* test)
- One-Sample Test about Median (Wilcoxon Signed Rank Sum)
- One-Sample Test about Proportion (Binomial test)

Assumptions

- Type of data – Discrete Numerical
- Randomization – Data gathered randomly (iid)
- Population distribution – Each trial is Bernoulli
- Sample size – Used for any sample size

Hypotheses

Null hypothesis: The proportion is equal to some known value

$$H_0: \pi = \pi_0$$

Alternative hypothesis: The proportion is not equal to the known value in some way (depends on the research hypothesis)

$$H_a: \pi < \pi_0, \qquad H_a: \pi > \pi_0, \qquad H_a: \pi \neq \pi_0$$

Test Statistic

The test statistic is the total number of successes

$$x = \sum_{i=1}^{n} b_i$$

where $b_i = 0$ for a failure and $b_i = 1$ for a success.

Conclusion

Rejection region varies on the alternative hypothesis.

If $H_a: \pi < \pi_0$,

$$x \le \text{Binom}_{PPF,p}(\alpha)$$

If $H_a: \pi > \pi_0$,

$$\text{n} - x \le \text{Binom}_{PPF,1-p}(\alpha)$$

If $H_a: \pi \neq \pi_0$,

$$\sum_{i:P(X=i) \le P(X=x)} \text{Binom}_{PMF,p}(i) \le \alpha$$

Conclusion

If the observed value falls within the rejection region,

We reject H_0 at the α level of significance.

If the observed value falls outside the rejection region,

We fail to reject H_0 at the α level of significance.

P-value

Presuming that H_0 is true, the probability of observing a test statistic (X) more extreme than the one observed (x) is:

If $H_a\!: \mu < \mu_0$,

$$p = P(X \leq x) = \text{Binom}_{CDF,p}(x)$$

If $H_a\!: \mu > \mu_0$,

$$p = P(X \geq x) = \text{Binom}_{CDF,1-p}(n - x)$$

If $H_a\!: \mu \neq \mu_0$,

$$p = \sum_{i:P(X=i) \leq P(X=x)} \text{Binom}_{PMF,p}(i)$$

Recap

- Large-Sample Test about Mean/Proportion (*z* test)
- One-Sample Test about Mean (*t* test)
- One-Sample Test about Median (Wilcoxon Signed Rank Sum)
- One-Sample Test about Proportion (Binomial test)