IEOR 165 – Lecture 16
One-Sample Location Tests

1 Framework for One-Sample Location Tests

In many cases, we are interested in deciding if the mean (or median) of one group of random variables is equal to a particular constant. Even for this seemingly simple question, there are actually multiple different null hypothesis that are possible. And corresponding to each of these possible nulls, there is a separate *hypothesis test*. The term hypothesis test is used to refer to three objects:

- distribution for the null hypothesis;
- concept of extreme deviations within the null hypothesis;
- an equation to compute a p-value, using measured data, corresponding to the null.

The reason that there are multiple nulls is that we might have different assumptions about the distribution of the data, the variances, or the possible deviations.

A canonical example of a one-sample location test is the following null hypothesis

$$H_0 : X_i \sim \mathcal{N}(0, 1), \text{ where } X_i \text{ are iid.}$$

This is a simplistic example because we have assumed that the X_i are Gaussian random variables with a known variance, and the question we would like to decide is if the mean of the distribution is 0. In a more complex scenario, we might not know the variance of the distribution and would like to use the data to estimate the variance.

One of the most significant differences between various null hypothesis are the concepts of *one-tailed* and *two-tailed* tests. In a one-tailed test, the null hypothesis includes an assumption that deviations in only one-direction are possible. In a two-tailed test, the null hypothesis includes an assumption that deviations in both directions are possible. To make this more concrete, recall the coin-flipping example from the previous lecture: The null hypothesis was that the coin is not biased (towards heads or towards tails), and we measured 40 heads out of $n = 100$ trials. And we measured deviation by saying that extreme examples were the union of $0, \ldots, 40$ or $60, \ldots, 100$ heads. This is an example of a two-tailed test, because we considered deviations in both directions (very high and very low numbers of heads). However, suppose our null hypothesis is that the coin is not biased towards tails. Then in this case, we would say that extreme examples were only $0, \ldots, 40$ heads; this would have been an example of a one-sided test.

2 One-Sample Gaussian Tests

Suppose the null hypothesis involves a distribution where the data X_i are iid Gaussians with a known variance σ^2 (and standard deviation σ). There are three possible concepts of extreme deviations, each with a corresponding equation. Let $Z \sim \mathcal{N}(0,1)$, and suppose μ_0 is a known constant. Then the three tests are

- One-Tailed, where $H_0 : X_i \sim \mathcal{N}(\mu, \sigma^2),$ where X_i are iid and $\mu \geq \mu_0$

$$p = \mathbb{P}\left(Z < \sqrt{n} \cdot \left(\frac{\overline{X} - \mu_0}{\sigma}\right)\right)$$

- One-Tailed, where $H_0 : X_i \sim \mathcal{N}(\mu, \sigma^2),$ where X_i are iid and $\mu \leq \mu_0$

$$p = \mathbb{P}\left(Z > \sqrt{n} \cdot \left(\frac{\overline{X} - \mu_0}{\sigma}\right)\right)$$

- Two-Tailed, where $H_0 : X_i \sim \mathcal{N}(\mu_0, \sigma^2),$ where X_i are iid

$$p = \mathbb{P}\left(|Z| > \sqrt{n} \cdot \left|\frac{\overline{X} - \mu_0}{\sigma}\right|\right)$$

The intuition behind these tests is that, by the properties of iid Gaussian distributions, we know that $(\sigma/\sqrt{n}) \cdot Z + \mu$ has distribution $\mathcal{N}(\mu, \sigma^2/n)$. And so these tests are looking at the probability we would see something as (or more) extreme than $\overline{X}$, which by the properties of iid Gaussians has distribution $\mathcal{N}(\mu, \sigma^2/n)$. As for computation of the p-values, we can get these values using a calculator/computer or using a standard Z-table.

This test is sometimes used even if the distribution in the null hypothesis is not Gaussian. The reason is the central limit theorem says that the distribution of $\sqrt{n}(\overline{X} - \mathbb{E}(X_i))$ converges in distribution to that of $\mathcal{N}(0, \sigma^2)$. And so if we have a large amount of data ($n = 30$ is common rule of thumb, though the meaning of "large" depends on how skewed the distribution of the null hypothesis is) then we can approximate the p-value using this test.

2.1 Example: Chemical pH Testing

Q: In one chemical process, it is very important that a particular solution used as a reactant have a pH of exactly 8.15. A method for determining pH available for solutions of this type is known to give measurements that are normally distributed with a mean equal to the actual pH and a standard deviation of 0.02. We wish to use a test such that if the pH is actually equal to 8.15 then this conclusion will be reached with probability equal to 0.90. Suppose the sample size is 5. If $\overline{X} = 8.26$, what is the conclusion?

A: We first specify the null hypothesis. We have that

$$H_0 : X_i \sim \mathcal{N}(8.15, (0.02)^2), \text{ where } X_i \text{ are iid},$$

and we are interested in two-tailed deviations in this null hypothesis since we need the pH to be exactly 8.15. Let $Z \sim \mathcal{N}(0, 1)$. Then under H_0, we compute

$$\begin{aligned} p &= \mathbb{P}\Big(|Z| > \sqrt{n} \cdot \Big|\frac{\overline{X} - \mu_0}{\sigma}\Big|\Big) = \mathbb{P}\Big(|Z| > \sqrt{5} \cdot \Big|\frac{8.26 - 8.15}{0.02}\Big|\Big) \\ &= \mathbb{P}(|Z| > 5.5) = 2 \cdot (1 - \mathbb{P}(Z < 5.5)) \approx 2 \cdot (1 - 0.998) = 0.004. \end{aligned}$$

Since we want the test to accept the null hypothesis 90% of the time when it is true, this means we choose an $\alpha = 1 - 0.9 = 0.1$. Thus, we reject the null hypothesis since $p = 0.004 < \alpha = 0.1$.

2.2 Example: Toothpaste Effectiveness

Q: An advertisement for a new toothpaste claims that it reduces cavities of children in their cavity-prone years. Cavities per year for this age group are normal with mean 3 and standard deviation 1. A study of 2,500 children who used this toothpaste found an average of 2.95 cavities per child. Assume that the standard deviation of the number of cavities of a child using this new toothpaste remains equal to 1. Is this data strong enough, at the 5 percent level of significance, to establish the claim of the toothpaste advertisement?

A: We first specify the null hypothesis. We have that

$$H_0 : X_i \sim \mathcal{N}(\mu, 1), \text{ where } X_i \text{ are iid, where } \mu \geq 3.$$

This null hypothesis is that the new toothpaste has a mean of 3 or more cavities, or equivalently that the new toothpaste is either as good or worse than the previous toothpaste. This means we are interested in a one-tailed test. We compute

$$\begin{aligned} p &= \mathbb{P}\Big(Z < \sqrt{n} \cdot \Big(\frac{\overline{X} - \mu_0}{\sigma}\Big)\Big) = \mathbb{P}\Big(Z < \sqrt{2500} \cdot \Big(\frac{2.95 - 3}{1}\Big)\Big) \\ &= \mathbb{P}(Z < -2.5) = 1 - \mathbb{P}(Z < 2.5) = 1 - 0.9938 = 0.006. \end{aligned}$$

We should reject the null hypothesis since $p = 0.006 < \alpha = 0.05$, and we conclude that the new toothpaste is an improvement over the previous one.

3 One-Sample t-Test

The Student's t distribution is the distribution of the random variable t defined as

$$t = \frac{Z}{\sqrt{V/\nu}},$$

where $Z \sim \mathcal{N}(0,1)$, $V \sim \chi^2_\nu$ has a chi-squared distribution with ν degrees of freedom, and Z, V are independent.

Now suppose the null hypothesis involves a distribution where the data X_i are iid Gaussians with *unknown* variance σ^2 (and standard deviation σ). There are three possible concepts of extreme deviations, each with a corresponding equation. Suppose μ_0 is a known constant. If t is a Student's t distribution with $n-1$ degrees of freedom, and

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2,$$

then the three tests are

- One-Tailed, where $H_0 : X_i \sim \mathcal{N}(\mu, \sigma^2),$ where X_i are iid and $\mu \geq \mu_0$

$$p = \mathbb{P}\left(t < \left(\frac{\overline{X} - \mu_0}{s/\sqrt{n}}\right)\right)$$

- One-Tailed, where $H_0 : X_i \sim \mathcal{N}(\mu, \sigma^2),$ where X_i are iid and $\mu \leq \mu_0$

$$p = \mathbb{P}\left(t > \left(\frac{\overline{X} - \mu_0}{s/\sqrt{n}}\right)\right)$$

- Two-Tailed, where $H_0 : X_i \sim \mathcal{N}(\mu_0, \sigma^2),$ where X_i are iid

$$p = \mathbb{P}\left(|t| > \left|\frac{\overline{X} - \mu_0}{s/\sqrt{n}}\right|\right)$$

The intuition behind these tests is that we have $n-1$ degrees of freedom because we are using the sample mean $\overline{X}$ in our estimate of the sample variance s^2. As for computation of the p-values, we can get these values using a calculator/computer or using a standard t-table.

This test is sometimes used even if the distribution in the null hypothesis is not Gaussian. The reason is the central limit theorem says that the distribution of $\sqrt{n}(\overline{X} - \mathbb{E}(X_i))$ converges in distribution to that of $\mathcal{N}(0, \sigma^2)$, and the weak law of large numbers says that s^2 converges in probability to σ^2. And so Slutsky's theorem implies that $(\overline{X} - \mu)/(s/\sqrt{n})$ converges in distribution to $\mathcal{N}(0,1)$. So if we have a large amount of data ($n = 30$ is common rule of thumb, though the meaning of "large" depends on how skewed the distribution of the null hypothesis is) then we can approximate the p-value using this test.

3.1 Example: Diet and Lifespan

Q: The following data relate to the ages at death of a certain species of rats that were fed 1 of 3 types of diets. Thirty rats of a type having a short life span of an average of 17.9 months were randomly divided into 3 groups of 10 each. The sample means and variances of ages at death (in months) are:

	Very Low Calorie	Moderate Calorie	High Calorie
Sample mean	22.4	16.8	13.7
Sample variance	24.0	23.2	17.1

Use a t-test to study whether a very low calorie diet leads to an increase in life-span, using a significance level of 0.01.

A: We first specify the null hypothesis. We have that

$$H_0 : X_i \sim \mathcal{N}(\mu, \sigma^2), \text{ where } X_i \text{ are iid, where } \mu \leq 17.9.$$

This null hypothesis is that the low calorie diet mice have a life span of 17.9 months or less. This means we are interested in a one-tailed t-test with 9-degrees of freedom (dof). We compute

$$\begin{aligned} p &= \mathbb{P}\Big(t > \Big(\frac{\overline{X} - \mu_0}{s/\sqrt{n}}\Big)\Big) = \mathbb{P}\Big(t > \Big(\frac{22.4 - 17.9}{\sqrt{24}/\sqrt{10}}\Big)\Big) \\ &= \mathbb{P}(t > 2.91) < 0.01. \end{aligned}$$

We should reject the null hypothesis since $p < \alpha = 0.01$, and we conclude that a very low calorie diet leads to an increase in life-span.

4 One-Sample Sign Test

In many cases, we may have a small sample size and the distribution is not Gaussian. In such instances, we cannot use the previous hypothesis tests. The sign test is an example of a test that is broadly applicable to settings without Gaussian distributed data. There are three possible concepts of extreme deviations, each with a corresponding equation. If W is the number of measurements that are larger than m_0, then the three tests are

- One-Tailed, where $H_0 : \text{median}(X_i) = m, \text{ where } X_i \text{ are iid and } m \geq m_0$

$$p_1 = \sum_{k \in \{0,\ldots,W\}} \binom{n}{k} 0.5^k \cdot 0.5^{n-k}$$

- One-Tailed, where $H_0 : \text{median}(X_i) = m, \text{ where } X_i \text{ are iid and } m \leq m_0$

$$p_2 = \sum_{k \in \{W,\ldots,n\}} \binom{n}{k} 0.5^k \cdot 0.5^{n-k}$$

- Two-Tailed, where H_0 : median$(X_i) = m_0,$ where X_i are iid

$$p_3 = 2\min\{p_1, p_2\}$$

The intuition behind these tests is that W follows a binomial distribution.

This is an example of what is known as a *nonparametric* test. The reason for this name is that the null hypothesis is fairly broad, and it does not depend upon the particular distribution. This is in contrast to say the t-test, in which the null hypothesis specifies that the data is generated by a Gaussian distribution with known mean but an unknown variance parameter. The previous hypothesis tests we considered are *parametric* tests because they depend on a small number of parameters.