

Testing hypotheses about the mean

Earlier, we saw how to test hypotheses about a proportion, using properties of the Binomial distribution. It is quite tricky.

Testing hypotheses is actually easier when the underlying distribution is continuous, because then we can choose the rejection region to have any specified significance level.

First of all we will revisit the “snowy day” example, and work through the steps given at the end of the chapter on testing hypotheses about a proportion.

Example The time required by workers to complete an assembly job usually has a mean of 50 minutes and a standard deviation of 8 minutes. On a snowy day, the supervisor wonders if they are working at the same speed as usual. In fact, he suspects that they have got slower and wants to find out if he is right. He intends to record the times that 60 workers take to complete one assembly job apiece.

Time in minutes to complete job:	mean μ , s.d. 8
Question:	μ used to be 50, have they got slower?
Unknown parameter:	μ
Two hypotheses:	$\mu \leq 50$ and $\mu > 50$
Null hypothesis:	$\mu \leq 50$
Trial:	record time taken by 60 workers
Test statistic:	$\bar{X}$ = mean of these 60 times
Rejection region:	$[c, \infty)$ for some $c > 50$
Test	$\bar{X} \geq c$
Probability model:	$\bar{X} \sim N\left(\mu, \frac{8^2}{60}\right)$
Tolerance limit:	$\alpha \leq 5\%$

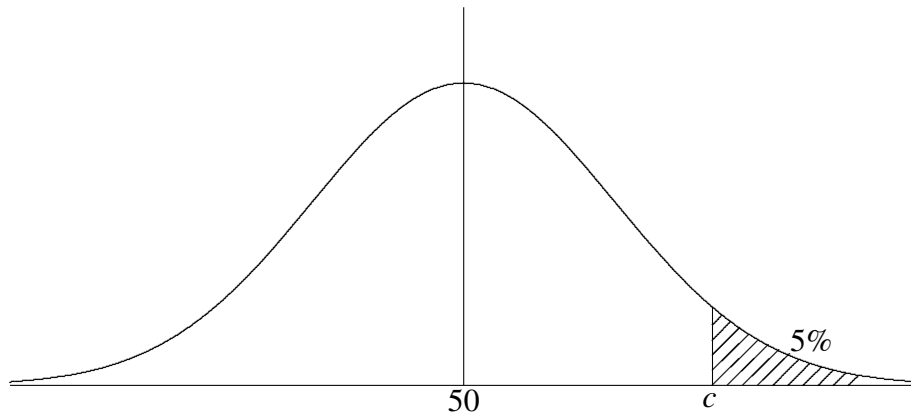
The extreme value of μ satisfying H_0 is $\mu = 50$. If $\mu = 50$ then $\mathbb{P}(\text{Type I error}) =$

$$\mathbb{P}(\bar{X} \geq c) = \mathbb{P}\left(\frac{\bar{X} - 50}{8/\sqrt{60}} \geq \frac{c - 50}{8/\sqrt{60}}\right) = \mathbb{P}\left(Z \geq \frac{c - 50}{8/\sqrt{60}}\right),$$

where $Z \sim N(0, 1)$. We want this probability to be 0.05, so

$$\frac{c - 50}{8/\sqrt{60}} = z_{0.05} = 1.645,$$

so $c = 1.645 \times 8/\sqrt{60} + 50 = 51.70$.



Therefore the rejection region is $[51.70, \infty)$, and the significance level is exactly 5%, because the distribution is continuous.

Suppose that the supervisor observed $\bar{x} = 53$. Now, $53 > 51.70$, so 53 is in the rejection region, so H_0 is rejected at the 5% significance level. He concludes that the workers have got slower.

The P-value is $\mathbb{P}(\bar{X} \geq 53)$ if $\mu = 50$. This is equal to

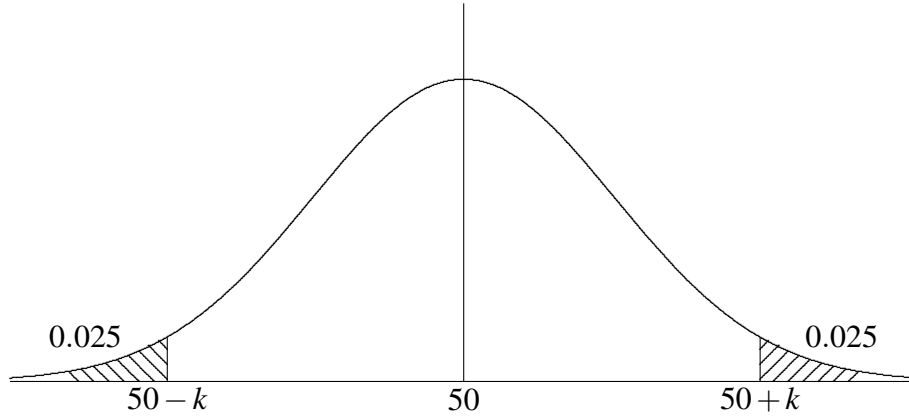
$$\mathbb{P}\left(Z \geq \frac{53 - 50}{8/\sqrt{60}}\right) = \mathbb{P}(Z \geq 2.90) = 0.0019.$$

This is very small, so the supervisor should be worried! If the workers have not got slower, only two samples per thousand would give a mean time as large as this.

Instead, suppose that the supervisor had asked “Has the mean time changed?” Then we would have

Null hypothesis:	$\mu = 50$
Alternative hypothesis:	$\mu \neq 50$
Rejection region:	$(-\infty, c_1] \cup [c_2, \infty)$ for some real numbers c_1 and c_2 with $c_1 < 50 < c_2$.

The normal distribution is symmetric, so it is sensible to put $c_2 = 50 + k$ and $c_1 = 50 - k$ for some k . Assume that we still want the significance level to be 5%.



If $\mu = 50$, we need $\mathbb{P}(\bar{X} \geq 50 + k) = \mathbb{P}(\bar{X} \leq 50 - k) = 0.025$. Therefore

$$\mathbb{P}\left(Z \geq \frac{k}{8/\sqrt{60}}\right) = 0.025,$$

and so

$$\frac{k}{8/\sqrt{60}} = z_{0.025} = 1.96.$$

Therefore $k = 1.96 \times 8/\sqrt{60} = 2.02$, and so the rejection region is

$$(-\infty, 47.98] \cup [52.02, \infty)$$

If the supervisor observes $\bar{x} = 53$, this is still in the rejection region, so H_0 is rejected at the 5% significance level.

Now “more extreme” values are those further from 50 than $\bar{x}$, in either direction, so the P-value is the value of $\mathbb{P}(\bar{X} \geq 53 \text{ or } \bar{X} \leq 47)$ if $\mu = 50$. This is equal to $2\mathbb{P}(\bar{X} \geq 53) = 0.0038$.

Notice that tests and their conclusions depend on the questions being asked, and therefore on the hypotheses.

In that example, we took $\bar{X}$ to be the test statistic and had to calculate the rejection region. In fact, it would have been easier to take the test statistic to be

$$\frac{\bar{X} - 50}{8/\sqrt{60}}.$$

General procedure for hypothesis tests of the mean value

Suppose that μ is the unknown mean of some large population with variance σ^2 ; that H_0 and H_1 are phrased in terms of μ ; that $\bar{X}$ is the mean of a random sample of size n ; and that we can assume that $\bar{X}$ is (approximately) normal. Remember that $\mathbb{E}(\bar{X}) = \mu$ and $\text{Var}(\bar{X}) = \sigma^2/n$.

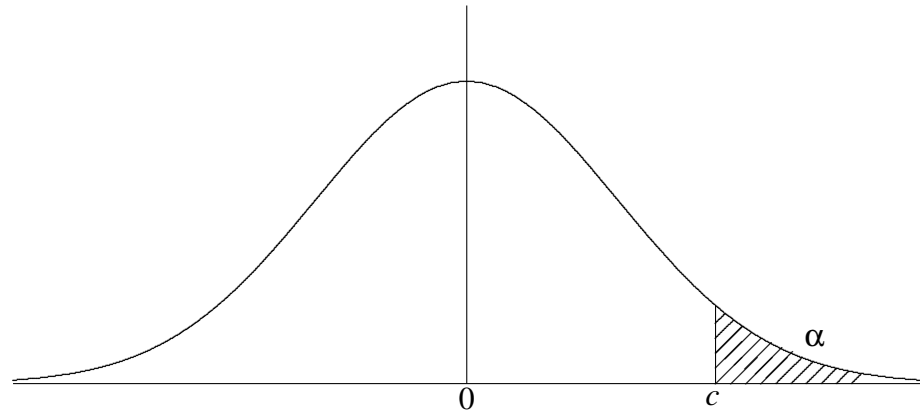
1. Population variance σ^2 known

(a) One-sided tests (also called one-tailed tests)

- (i) $H_0: \mu \leq \mu_0$, where μ_0 is known;
 $H_1: \mu > \mu_0$.

We know that $\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0, 1)$ under H_0 ,

so it is convenient to take $\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$ as the test statistic. For significance level α , the rejection region should be $[c, \infty)$ where $c = z_\alpha$.



Given the sample mean $\bar{x}$ of the data, reject H_0 if $\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \geq z_\alpha$.

- (ii) $H_0: \mu \geq \mu_0$, where μ_0 is known;
 $H_1: \mu < \mu_0$.

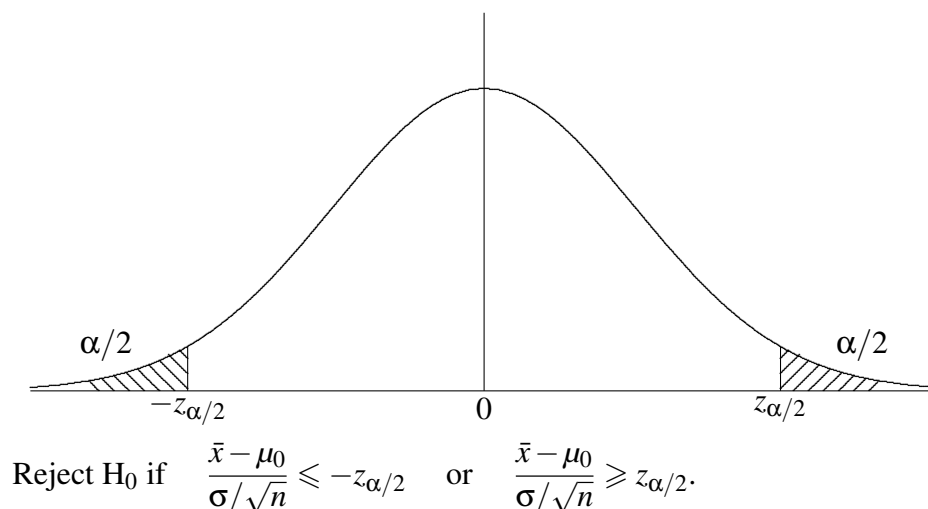
By a similar argument, reject H_0 if $\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \leq -z_\alpha$.

(b) Two-sided tests (also called two-tailed tests)

$H_0: \mu = \mu_0$, where μ_0 is known;

$H_1: \mu \neq \mu_0$.

Use the same test statistic $\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$ but with the symmetric rejection region $(-\infty, -z_{\alpha/2}] \cup [z_{\alpha/2}, \infty)$.



2. σ^2 unknown, but n large ($n \geq 30$)

As Case 1, but replace the known σ by the sample standard deviation s .

Thus the test statistic is

$$\frac{\bar{X} - \mu_0}{S/\sqrt{n}}$$

but the rejection region does not change.

These tests are all called “one-sample z -tests”.

Relationship between hypothesis tests and confidence intervals

If we do a two-sided z -test with $H_0: \mu = \mu_0$, test statistic $\bar{X}$ and significance level α ,

$$\begin{aligned} \bar{x} \in \text{Rejection Region} &\iff \bar{x} \leq \mu_0 - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \quad \text{or} \quad \bar{x} \geq \mu_0 + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \\ &\iff \mu_0 \geq \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \quad \text{or} \quad \mu_0 \leq \bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \\ &\iff \mu_0 \notin \left(\bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right) \\ &\iff \mu_0 \notin (1 - \alpha)\text{-confidence interval for } \mu. \end{aligned}$$

Tests of a proportion

The proportion p is just the mean of a Bernoulli distribution. The sample total has a Binomial distribution. If n is small then we construct rejection regions directly from tables of the Binomial distribution, as in the pizza example. If n is large, we use the normal approximation. If the true proportion is p_0 then the true variance of $\bar{X}$ is p_0q_0/n , where $q_0 = 1 - p_0$, so we use method 1 (“variance known”).

Tests of a Poisson mean

If $X_1, X_2, \dots, X_n$ are independent random variables with the Poisson distribution with mean λ , then the sample total $X_1 + X_2 + \dots + X_n \sim \text{Poisson}(n\lambda)$. If n is small, we construct rejection regions directly from tables of the Poisson distribution. If $n\lambda > 9$ then we use the normal approximation, which is justified by the Central Limit Theorem. If the true mean is λ_0 then the true variance of $\bar{X}$ is λ_0/n , so we use method 1 (“variance known”).

For the Binomial and Poisson distributions, if we know the mean then we know the variance. So there is no longer a simple relationship between two-sided hypothesis test and confidence intervals. For example, consider $\text{Bin}(n, p)$ when p is unknown, and suppose that p_0 corresponds to the null hypothesis and $\hat{p}$ is the point estimate from the data. The length of the confidence interval is proportional to

$$\sqrt{\frac{\hat{p}\hat{q}}{n}},$$

where $\hat{q} = 1 - \hat{p}$, but the length of the interval between the two parts of the rejection region is proportional to

$$\sqrt{\frac{p_0q_0}{n}},$$

where $q_0 = 1 - p_0$.