# Stat 401, section 8.1 Hypotheses and Test Procedures

notes by Tim Pilachowski

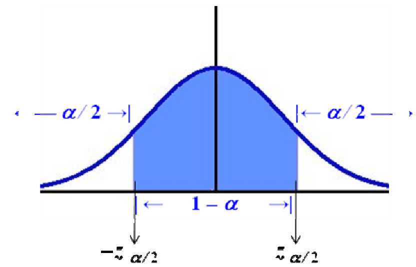
The process of chapter 8 will parallel that of chapter 7, but focus on hypothesis tests rather than confidence intervals: definition, theory and underlying concepts ($9^{th}$ and $8^{th}$ 8.1), large-sample situations ($9^{th}$ 8.2, $8^{th}$ 8.2a & 8.4), small-sample situations ($9^{th}$ 8.3, $8^{th}$ 8.2b & 8.4), and proportions ($9^{th}$ 8.4, $8^{th}$ 8.3 & 8.4). The course schedule contains a $9^{th}$–$8^{th}$ edition concordance.

The vocabulary and notation carry over, with some new additions.

$\alpha$ = significance level = probability of a Type I error

$\beta$ = probability of a Type II error

We'll take a closer look at these in a little while.




On this picture of a sampling distribution (two-sided scenario), the blue shaded area represents $100(1 - \alpha)\%$ of the possible sample statistics $(\hat{\theta}, \bar{x}, \hat{p})$ surrounding the population parameter $(\theta, \mu, p)$ in the middle.

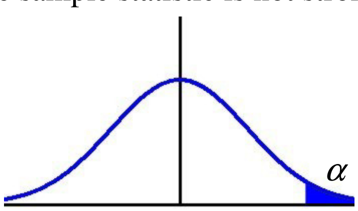
In both large-sample and small-sample-normal-population situations, the sampling distribution is symmetric, so each tail must contain $\alpha/2$.

The left and right boundaries will be the negative and positive values of $z$ that mark the lower and upper $100(\alpha/2)\%$.

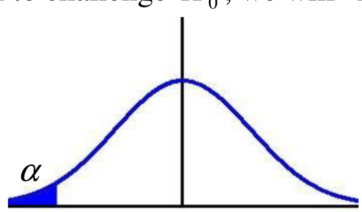
Here's the underlying concept of a hypothesis test. We'll assume we know something about a population parameter $\theta$. If we're correct, then the sample statistic $\hat{\theta}$ that we calculate has a high probability, $100(1 - \alpha)\%$, of being close to our hypothesized $\theta$. If, however, we get a sample statistic $\hat{\theta}$ which has a low probability of occurring, we'll question whether our initial assumption about population parameter $\theta$ was correct. (Think "proof by contradiction".)

Definitions: The initial assumption ("prior belief claim") is called the **null hypothesis**, and is denoted by $H_0$. The challenge to the null hypothesis is called the **alternate hypothesis** ("assertion that is contradictory to $H_0$") and is denoted by $H_a$.

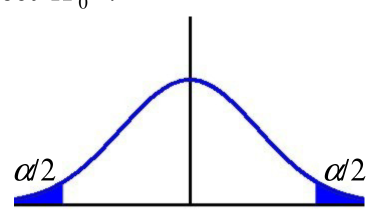
If the sample statistic presents strong enough evidence that $H_0$ is false, we will "reject $H_0$".
If the sample statistic is not strong enough to challenge $H_0$, we will "fail to reject $H_0$".




$H_0: \theta \le \theta_0$
$H_a: \theta > \theta_0$




$H_0: \theta \ge \theta_0$
$H_a: \theta < \theta_0$




$H_0: \theta = \theta_0$
$H_a: \theta \ne \theta_0$

$8^{th}$ edition approach: The blue-shaded area in each of the pictures above has traditionally been called the "rejection region". If a calculated test statistic fell into the rejection region, the decision would be to reject $H_0$. Otherwise the decision would be to fail to reject $H_0$.

[Side note: The text shows all null hypotheses as "$=$". In this class, we'll use "$\le$" and "$\ge$" for one-tailed tests.]

Example A - hypotheses. To be useful, ball bearings need to have a constant mean diameter of 0.50 cm; those much larger or smaller can wreak havoc. A sample of 50 has a mean diameter of 0.51 cm with $s$ = 0.04. What hypotheses should be tested, and why?

Example B - hypotheses. In their 1992 study of human internal body temperature, Mackowiak, Wasserman and Levine had a sample size of 148, and calculated $\bar{x}$ = 98.25 and $s$ = 0.73. What hypotheses should be tested, and why?

Example C - hypotheses. A current medical treatment has been effective for 70% of the patients to whom it was administered. In a clinical trial of a new treatment with 26 participants, 19 experienced reduction of symptoms. What hypotheses should be tested, and why?

Before we continue, we need to prove that our null-alternate-hypotheses approach provides what we need. Specifically, do the hypothesis tests described above have a significance level equal to $\alpha$?
We'll focus on a test of a population mean $\mu$, with the understanding that a proof involving another population parameter would proceed along similar lines.

For a two-tailed $z$-test, our hypotheses are $\begin{array}{l} H_0 : \mu = \mu_0 \\ H_a : \mu \neq \mu_0 \end{array}$.

Our decision rule would be "reject $H_0$ if either $\bar{x} \le \mu_0 - z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$ or $\bar{x} \ge \mu_0 + z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$.

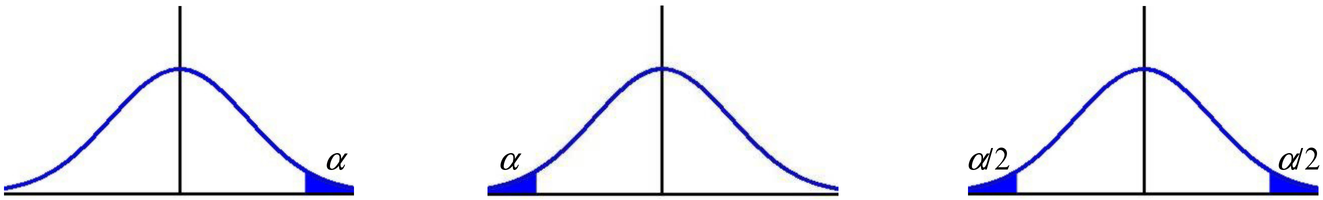
notes on the proof:

$$
\begin{aligned}
& P(\text{ reject } H_0 \text{ when } H_0 \text{ is correct }) \\
&= P\left( \bar{X} \le \mu_0 - z_{\alpha/2}\left(\tfrac{\sigma}{\sqrt{n}}\right) \text{ or } \bar{X} \ge \mu_0 + z_{\alpha/2}\left(\tfrac{\sigma}{\sqrt{n}}\right) \text{ when } \mu = \mu_0 \right) \\
&= P\left( \bar{X} - \mu_0 \le - z_{\alpha/2}\left(\tfrac{\sigma}{\sqrt{n}}\right) \text{ or } \bar{X} - \mu_0 \ge z_{\alpha/2}\left(\tfrac{\sigma}{\sqrt{n}}\right) \text{ when } \mu = \mu_0 \right) \\
&= P\left( \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \le - z_{\alpha/2} \text{ or } \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \ge z_{\alpha/2} \text{ when } \mu = \mu_0 \right) \\
&= P\left( \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \le - z_{\alpha/2} \text{ or } \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \ge z_{\alpha/2} \right) \\
&= P\left( Z \le - z_{\alpha/2} \text{ or } Z \ge z_{\alpha/2} \right) \\
&= \frac{\alpha}{2} + \frac{\alpha}{2} \\
&= \alpha
\end{aligned}
$$

The proof of the lower-tailed $z$-test would be similar, using "<", "–" and $z_\alpha$.
The proof of the upper-tailed $z$-test would be similar, using ">", "+" and $z_\alpha$.

In practice, when we use $S$ as a point estimate for $\sigma$, we rely on the CLT: When $n$ is large enough, the random variable $Z = \frac{\overline{X}-\mu}{S/\sqrt{n}}$ has approximately a standard normal distribution.

While the traditional method of determining the decision rule for a hypothesis test involves specifying a rejection region (as is done in the 8th edition of your text), in current practice it is much more common to calculate a $p$-value (as is done in section 8.4 of the 8th edition and throughout chapter 8 in the 9th edition).
"The ***P*-value** is the probability, calculated assuming that the null hypothesis is true, of obtaining a value of the test statistic at least as contradictory to $H_0$ as the value calculated from the available sample data."




| | | |
|---|---|---|
| $H_0: \theta \le \theta_0$ | $H_0: \theta \ge \theta_0$ | $H_0: \theta = \theta_0$ |
| $H_a: \theta > \theta_0$ | $H_a: \theta < \theta_0$ | $H_a: \theta \ne \theta_0$ |

If the $P$-value $\le \alpha$, we will reject $H_0$. Otherwise we will fail to reject $H_0$.
[Side note: The text shows all null hypotheses as "$=$". In this class, we'll use "$\le$" and "$\ge$" for one-tailed tests.]

Now we need to prove that the $P$-value provides what we need.
Specifically, does the $P$-value decision rule have a significance level equal to $\alpha$?

For a two-tailed $z$-test, our hypotheses are $\begin{matrix} H_0: \theta = \theta_0 \\ H_a: \theta \ne \theta_0 \end{matrix}$.

Our decision rule is "reject $H_0$ if either $z \le -z_{\alpha/2}$ or $z \ge z_{\alpha/2}$, which we condense to $|z| \ge z_{\alpha/2}$.
Theorem: The $p$-value of a two-sided $z$-test is a function of $z$ alone, and moreover $p = p(z) = 2(1-\Phi(|z|))$.

notes on the proof:

$$|z| \ge z_{\alpha/2}$$
$$\Phi(|z|) \ge \Phi(z_{\alpha/2})$$
$$\Phi(|z|) \ge 1-\frac{\alpha}{2}$$
$$\frac{\alpha}{2} \ge 1-\Phi(|z|)$$
$$\alpha \ge 2(1-\Phi(|z|))$$

That is, $2(1-\Phi(|z|))$ is the smallest of the set of $\alpha$s for which $H_0$ will be rejected.

For the two-tailed hypothesis tests, we must multiply by 2, since $\alpha$ was split into two tails.

For each of the one-tailed tests, $p = p(z) = 1-\Phi(|z|) = \Phi(-|z|)$, since all of $\alpha$ is in one tail.

The proof of the lower-tailed $z$-test would be similar, using $z \le -z_\alpha$.
The proof of the upper-tailed $z$-test would be similar, using $z \ge z_\alpha$.

Last piece of today's Lecture: Errors in Hypothesis Testing.

| | Actual: $H_0$ is True | Actual: $H_0$ is False |
|---|---|---|
| Decision: Reject $H_0$ | | |
| Decision: Fail to Reject $H_0$ | | |

$\alpha$ = significance level = probability of a Type I error

$\beta$ = probability of a Type II error

While there will be only one value for $\alpha$, there will be many possible values for $\beta$. There are many possible actual values of the population parameter $\theta$ for which we would (incorrectly) fail to reject $H_0$.

Example A - errors. To be useful, ball bearings need to have a constant mean diameter of 0.50 cm; those much larger or smaller can wreak havoc. A sample of 50 has a mean diameter of 0.51 cm with $s = 0.04$. In this context, what are the Type I and Type II errors? In this context, which would be more serious?

Example B - errors. In their 1992 study of human internal body temperature, Mackowiak, Wasserman and Levine had a sample size of 148, and calculated $\bar{x}$ = 98.25 and $s$ = 0.73. In this context, what are the Type I and Type II errors? In this context, which would be more serious?

Example C - errors. A current medical treatment has been effective for 70% of the patients to whom it was administered. In a trial of 26 patients, 19 experienced reduction of symptoms. In this context, what are the Type I and Type II errors? In this context, which would be more serious?