Underlying reasoning

1. Assume some hypothesis about the parameter of interest is true (call it the `null' or H_0)

2. Sample from a population to get an estimate of that parameter.

3. Determine how likely it was to have obtained that estimate (or an estimate more extreme), under the condition that the null hypothesis is true. (Note: `extremeness' is determined by the alternative hypothesis)

4. If the probability of obtaining that estimate (or one more extreme) is *small* under H_0, then we **assume** that it is more likely that the null hypothesis is incorrect than it is to have observed such an extreme estimate. Hence we "reject H_0".

5. If the probability of obtaining that estimate (or one more extreme) is *not small* under H_0, then the null hypothesis can plausibly explain the observed estimate. Hence we "Fail to reject H_0".

Conducting a hypothesis test (7 steps)

1. Write down the probability model and given information.

2. Determine the Null (H_0) and Alternative (H_A) hypothesis concerning the parameter of interest.

3. Calculate a test statistic (T^*) whose distribution, say Q, does not depend on any unknown quantities. (necessary to calculate accurate probability statements about T^*)

4. Determine the rejection regions under the distribution of Q, based on the type one error probability α. (i.e., $>Q_{\alpha/2}$ and $<-Q_{\alpha/2}$)

5. Check to see if the test statistic lies in the rejection regions. (i.e., is $|T^*|>Q_{\alpha/2}$?)

6. If the test statistic is in the rejection region then "reject H_0", otherwise "Fail to reject H_0".

7. Write your conclusions in a simple English sentence.

A note on the form of the test statistic T^*

The test statistic, T^*, may take any ugly form or shape, as long as the distribution of T^* , say Q, does not depend on any unknown quantity.

There are often many different T^*'s, and sometimes there are even several `correct' T^*'s for a given problem. The most commonly used test statistics, however, all have the following similar form:

$$T^* = \frac{est - E[est \mid H_0]}{\sqrt{Var[est \mid H_0]}} \sim Q$$

The wisdom of using this form is that Q will always be approximately normal is large sample, regardless of the underlying situation (that is, as long as the CLT and LLN hold).

This is why we use $T^* = \frac{\sqrt{n}\left(\overline{X} - \mu_0\right)}{\sigma}$ (here Q~N(0,1)) instead of $T^* = \overline{X}$ (here Q~N(μ_0, σ^2)) to test the hypothesis that H_0: $\mu=\mu_0$.

One-sided tests

We know that the rejection regions for the hypothesis test H_0: $\mu=\mu_0$ versus H_A: $\mu\neq\mu_0$ are the two tail areas $<-Q_{\alpha/2}$ and $>Q_{\alpha/2}$.

For obvious reasons, this is called a `two-sided' test. Likewise there are also two one-sided tests, where the entire rejection region, α, is one tail:

For testing H_0: $\mu=\mu_0$ versus H_A: $\mu>\mu_0$ we reject when $T^* > Q_\alpha$

For testing H_0: $\mu=\mu_0$ versus H_A: $\mu<\mu_0$ we reject when $T^* < -Q_\alpha$

Anatomy of a hypothesis test

H_0 and H_A are the null and alternative hypotheses

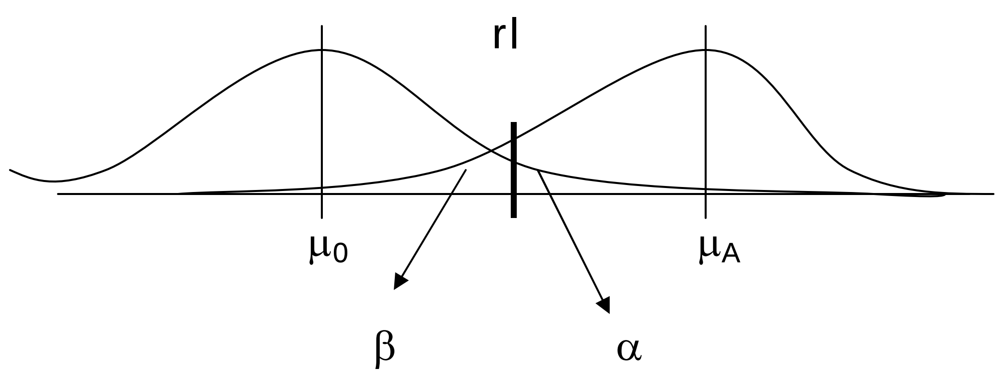
T^* is the test statistic (calculated assuming H_0 is true)

α = P(reject H_0 | H_0 true) is the type one error
β = P(Fail to reject H_0 | H_A true) is the type II error

Example

Model $X_1,\ldots,X_n \sim N(\mu, \sigma^2)$ (σ^2 known).

H_0: $\mu=\mu_0$ versus H_A: $\mu=\mu_A$ ($\mu_0<\mu_A$ are simple hypotheses)



Notice that both errors are conditional probabilities.
The rejection line (rl) depends only on H_0
Power is defined as 1-β. What happens as μ_A moves?

Measuring evidence against a null hypothesis

A hypothesis test tells you what you should do, i.e. reject or accept H_0.

To measure the evidence *against* the null hypothesis we need another tool. This tool is called the p-value.

p-value = $P(|Q|> T_{obs}^{*} | H_0)$ for a two sided test

$= P(Q< T_{obs}^{*} | H_0)$ for a one sided
$= P(Q> T_{obs}^{*} | H_0)$ for a one sided

where T_{obs}^{*} is the observed test statistic (remember we know the distribution of Q under H_0).

The p-value is just the smallest type one error that would have rejected H_0. It is the probability of observing a test statistic that extreme or more so under the null hypothesis.

Example 1

Suppose I wish to determine if Brown students tend to have higher IQ's than students at Harvard. I know that the distribution of IQ's in the population is normal. Furthermore, it comes to me in a dream that the variance of each population is 15^2.

1. Probability model

$X_1,\ldots,X_{n=200} \sim N(\mu_b,\sigma^2)$ ($\sigma^2 = 15^2$)
represents Brown IQ's with $\bar{X} = 120$

$Y_1,\ldots,Y_{m=200} \sim N(\mu_h,\sigma^2)$ ($\sigma^2 = 15^2$)
represents Harvard IQ's with $\bar{Y} = 115$

2. Construct hypotheses

H_0: $\mu_b - \mu_h \le 0$ versus H_A: $\mu_b - \mu_h > 0$

3. Calculate test statistic

$$T^* = \frac{\bar{X} - \bar{Y} - 0}{\sqrt{\frac{\sigma_b^2}{n} + \frac{\sigma_h^2}{m}}} = \frac{120 - 115}{\sqrt{\frac{15^2}{200} + \frac{15^2}{200}}} = \frac{5}{1.5} = 3.33$$

4. Distribution of Q?

under H_0, the distribution of T^* is $N(0,1) = Q$

5-6. Where does T^* lie?

Because $T_{obs}^* = 3.33 > 1.645 = Z_\alpha = Q_\alpha$ when $\alpha = 0.05$, we reject H_0 at the 5% level.

7. Conclusions

These data do not support the hypothesis that students at Harvard tend to have IQ's at least as great at Brown student IQ's, as the 5% significance level.

(Add confidence interval here!!)

The p-value

p-value = $P(\ Q > T_{obs}^* \mid H_0)$

p-value $= P(\ Z > 3.33) = 0.0012$.

Example 2

1. Probability model

$X_1,\ldots,X_{n=25} \sim N(\mu_b,\sigma^2)$ ($\sigma^2 = 36$) and $\bar{X} = 10$

2. Construct hypotheses

H_0: $\mu=12$ versus H_A: $\mu \neq 12$

3. Calculate test statistic

$$T^* = \frac{\sqrt{n}\left(\bar{X} - \mu_0\right)}{\sigma} = \frac{\sqrt{25}(10-12)}{6} = -1.67$$

4. Distribution of Q?

under H_0, the distribution of T^* is $N(0,1)=Q$

5-6. Where does T^* lie?

$T_{obs}^* = |-1.67\ | < 1.96 = Z_{\alpha/2}$ when $\alpha=0.05$ and
$T_{obs}^* = |-1.67\ | > 1.645 = Z_{\alpha/2}$ when $\alpha=0.10$

7. Conclusions

reject H_0 at the 10% level, but not at the 5% level.

p-value $=P(\ |Z| > -1.67\ ) = 2(0.047) = 0.094$

Example 2a - one sided version

1. Probability model

$X_1,\ldots,X_{n=25} \sim N(\mu_b,\sigma^2)$ ($\sigma^2 = 36$) and $\bar{X} = 10$

2. Construct hypotheses

H_0: $\mu \geq 12$ versus H_A: $\mu < 12$

3. Calculate test statistic

$$T^* = \frac{\sqrt{n}\left(\bar{X} - \mu_0\right)}{\sigma} = \frac{\sqrt{25}(10-12)}{6} = -1.67$$

4. Distribution of Q?

under H_0, the distribution of T^* is N(0,1)=Q

5-6. Where does T^* lie?

$T_{obs}^* = -1.67 < -1.96 = -Z_\alpha$ when $\alpha=0.05$ and
$T_{obs}^* = -1.67 < -1.28 = -Z_\alpha$ when $\alpha=0.10$

7. Conclusions

reject H_0 at the 5% level and the 10% level.

p-value =P(Z < -1.67) = 0.047

Example 2a - one sided version

1. Probability model

$X_1,\ldots,X_{n=25} \sim N(\mu_b,\sigma^2)$ (σ^2 = 36) and $\bar{X} = 10$

2. Construct hypotheses

H_0: $\mu \leq 12$ versus H_A: $\mu > 12$

3. Calculate test statistic

$$T^* = \frac{\sqrt{n}\left(\bar{X} - \mu_0\right)}{\sigma} = \frac{\sqrt{25}(10-12)}{6} = -1.67$$

4. Distribution of Q?

under H_0, the distribution of T^* is N(0,1)=Q

5-6. Where does T^* lie?

$T_{obs}^* = -1.67 \not> 1.96 = Z_\alpha$ when α=0.05 and
$T_{obs}^* = -1.67 \not> -1.28 = Z_\alpha$ when α=0.10

7. Conclusions

Fail to reject H_0 at the 5% level and the 10% level.

p-value =P(Z > -1.67) = 1-0.047 =0.953

Listing of test statistics:

One sample

Case 1

$X_1,\ldots,X_n \sim N(\mu,\sigma^2)$ (σ^2 known)

H_0: $\mu=\mu_0$

$$T^* = \frac{\bar{X}-\mu_0}{\sqrt{\frac{\sigma^2}{n}}} = \frac{\sqrt{n}\left(\bar{X}-\mu_0\right)}{\sigma} \sim N(0,1) \text{ under } H_0$$

Case 2

$X_1,\ldots,X_n \sim N(\mu,\sigma^2)$ (σ^2 unknown)

H_0: $\mu=\mu_0$

$$T^* = \frac{\sqrt{n}\left(\bar{X}-\mu_0\right)}{S} \sim t \text{ with df=n-1 under } H_0$$

Case 3

$X_1,\ldots,X_n \sim \text{Ber}(\theta)$ $(\hat{\theta} = \overline{X})$

H_0: $\theta=\theta_0$

$$T^* = \frac{\sqrt{n}\left(\hat{\theta}-\theta_0\right)}{\sqrt{\hat{\theta}\left(1-\hat{\theta}\right)}} \sim N(0,1) \text{ under } H_0$$

or

$$T^* = \frac{\sqrt{n}\left(\hat{\theta}-\theta_0\right)}{\sqrt{\theta_0\left(1-\theta_0\right)}} \sim N(0,1) \text{ under } H_0$$

Two sample

Case 1

$X_1,\ldots,X_n \sim N(\mu_X, \sigma_X^2)$ (σ_X^2 known)
$Y_1,\ldots,Y_m \sim N(\mu_Y, \sigma_Y^2)$ (σ_Y^2 known)

H_0: $\mu_X = \mu_Y$ or H_0: $\mu_X-\mu_Y = 0$

$$T^* = \frac{\overline{X}-\overline{Y}-\left(\mu_X-\mu_Y\right)}{\sqrt{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}}} \sim N(0,1) \text{ under } H_0$$

Case 2

$X_1,\ldots,X_n \sim N(\mu_X, \sigma_X^2)$ (σ_X^2 unknown)
$Y_1,\ldots,Y_m \sim N(\mu_Y, \sigma_Y^2)$ (σ_Y^2 unknown)

H_0: $\mu_X = \mu_Y$ or H_0: $\mu_X - \mu_Y = 0$ but assume $\sigma_X^2 = \sigma_Y^2$

$$T^* = \frac{\overline{X} - \overline{Y} - (\mu_X - \mu_Y)}{\sqrt{S_p^2\left(\frac{1}{n} + \frac{1}{m}\right)}} \sim \text{ t with df=m+n-2 under } H_0$$

where $S_p^2 = \frac{(n-1)S_X^2 + (m-1)S_Y^2}{n+m-2}$ is the `pooled' variance.

Case 3

$X_1,\ldots,X_n \sim N(\mu_X, \sigma_X^2)$ (σ_X^2 unknown)
$Y_1,\ldots,Y_m \sim N(\mu_Y, \sigma_Y^2)$ (σ_Y^2 unknown)

H_0: $\mu_X = \mu_Y$ or H_0: $\mu_X - \mu_Y = 0$ but $\sigma_X^2 \neq \sigma_Y^2$

$$T^* = \frac{\overline{X} - \overline{Y} - (\mu_X - \mu_Y)}{\sqrt{\frac{S_X^2}{n} + \frac{S_Y^2}{m}}} \sim Q \text{ under } H_0$$

In large samples Q~N(0,1).
In small samples use Q~t with df=min(n-1,m-1) or df=satterwaites correction.

Case 4

$X_1,\ldots,X \sim \text{Ber}(\theta_X)$ $\quad (\hat{\theta}_X = \overline{X})$
$Y_1,\ldots,Y \sim \text{Ber}(\theta_Y)$ $\quad (\hat{\theta}_Y = \overline{Y})$

H_0: $\theta_X = \theta_Y$ or H_0: $\theta_X - \theta_Y = 0$

$$T^* = \frac{\hat{\theta}_X - \hat{\theta}_Y - (\theta_X - \theta_Y)}{\sqrt{\frac{\hat{\theta}_X(1-\hat{\theta}_X)}{n} + \frac{\hat{\theta}_Y(1-\hat{\theta}_Y)}{m}}} \sim Z \text{ in large samples}$$

Power

Suppose I want to know how often this test will reject H_0: $\mu_0 = 12$ when the true mean is $\mu = 10$. Suppose also that I can only afford to select 25 subjects.

Model
$X_1,\ldots,X_{n=25} \sim N(\mu,36)$ and $\alpha=0.1$

Hypothesis
H_0: $\mu_0 = 12$ versus H_A: $\mu_A = 10$

power = $1-\beta$ $=1-P(\text{fail to reject } H_0 \mid H_0 \text{ false})$
$=P(\text{reject reject } H_0 \mid H_0 \text{ false})$
$= P(\text{reject reject } H_0 \mid H_A \text{ true})$

Whether or not we reject the null hypothesis depends on the form of the alternative hypothesis. Hence we have the following three cases in which to calculate the power:

Case 1: H_A: $\mu_A < 12$
Case 2: H_A: $\mu_A > 12$
Case 3: H_A: $\mu_A \neq 12$

Case 1: H_A: μ_A < 12

In this case, we reject when $T_{obs}^* < -Z_\alpha = -1.645$ ($\alpha=0.1$)

power = $1-\beta$ = P(reject H_0 | H_A true)

$$P\left(T^* < -Z_\alpha \mid H_A\right) = P\left(T^* < -1.645 \mid \mu = 10\right)$$
$$= P\left(\frac{\sqrt{n}\left(\overline{X}-\mu_0\right)}{\sigma} < -1.645 \mid \mu = 10\right)$$
$$= P\left(\overline{X}-\mu_0 < -1.645\frac{\sigma}{\sqrt{n}} \mid \mu = 10\right)$$
$$= P\left(\overline{X} < \mu_0 - 1.645\frac{\sigma}{\sqrt{n}} \mid \mu = 10\right)$$

To calculate we need to standardize:

$$1-\beta = P\left(\overline{X} < \mu_0 - 1.645\frac{\sigma}{\sqrt{n}} \mid \mu = 10\right)$$
$$= P\left(\frac{\sqrt{n}\left(\overline{X}-\mu_A\right)}{\sigma} < \frac{\sqrt{n}\left(\mu_0 - 1.645\frac{\sigma}{\sqrt{n}} - \mu_A\right)}{\sigma} \mid \mu = 10\right)$$
$$= P\left(Z < \frac{\sqrt{n}\left(\mu_0-\mu_A\right)}{\sigma} - 1.645 \mid \mu = 10\right)$$
$$= P\left(Z < \frac{5(12-10)}{6} - 1.645\right) = P\left(Z < 0.0217\right) = 1 - 0.492 = 0.5180$$

Case 2: H_A: $\mu_A > 12$

In this case, we reject when $T_{obs}^* >$ Z_α = 1.645 (α=0.1)

power = 1-β = P(reject H_0 | H_A true)

$$P\left(T^* > Z_\alpha \mid H_A\right) = P\left(T^* > 1.645 \mid \mu = 10\right)$$
$$= P\left(\frac{\sqrt{n}\left(\overline{X} - \mu_0\right)}{\sigma} > 1.645 \mid \mu = 10\right)$$
$$= P\left(\overline{X} < \mu_0 + 1.645\frac{\sigma}{\sqrt{n}} \mid \mu = 10\right)$$

To calculate we need to standardize:

$$1 - \beta = P\left(\overline{X} > \mu_0 + 1.645\frac{\sigma}{\sqrt{n}} \mid \mu = 10\right)$$
$$= P\left(\frac{\sqrt{n}\left(\overline{X} - \mu_A\right)}{\sigma} > \frac{\sqrt{n}\left(\mu_0 + 1.645\frac{\sigma}{\sqrt{n}} - \mu_A\right)}{\sigma} \mid \mu = 10\right)$$
$$= P\left(Z > \frac{\sqrt{n}\left(\mu_0 - \mu_A\right)}{\sigma} + 1.645 \mid \mu = 10\right)$$
$$= P\left(Z > \frac{5(12-10)}{6} + 1.645\right) = P\left(Z > 3.3117\right) \approx 0$$

There is almost no power. Does this make sense? Why? Draw a picture.

Case 3: H_A: $\mu_A \neq 12$

In this case, we reject when (α=0.1)

$T_{obs}^* > Z_{\alpha/2} = 1.96$ or $T_{obs}^* < -Z_{\alpha/2} = -1.96$

power = $1-\beta$ = P(reject H_0 | H_A true)

$$
\begin{aligned}
& P\left(\left|T^*\right| > Z_{\alpha/2} \mid H_A\right) \\
&= P\left(T^* < -Z_{\alpha/2} \text{ or } T^* > Z_{\alpha/2} \mid H_A\right) \\
&= P\left(T^* < -1.96 \mid \mu = 10\right) + P\left(T^* > 1.96 \mid \mu = 10\right) \\
&= P\left(\frac{\sqrt{n}\left(\bar{X} - \mu_0\right)}{\sigma} < -1.96 \mid \mu = 10\right) + P\left(\frac{\sqrt{n}\left(\bar{X} - \mu_0\right)}{\sigma} > 1.96 \mid \mu = 10\right) \\
&= P\left(\bar{X} < \mu_0 - 1.96\frac{\sigma}{\sqrt{n}} \mid \mu = 10\right) + P\left(\bar{X} > \mu_0 + 1.96\frac{\sigma}{\sqrt{n}} \mid \mu = 10\right) \\
&= P\left(\frac{\sqrt{n}\left(\bar{X} - \mu_A\right)}{\sigma} < \frac{\sqrt{n}\left(\mu_0 - \mu_A\right)}{\sigma} - 1.96 \mid \mu = 10\right) \\
&\qquad + P\left(\frac{\sqrt{n}\left(\bar{X} - \mu_A\right)}{\sigma} > \frac{\sqrt{n}\left(\mu_0 - \mu_A\right)}{\sigma} + 1.96 \mid \mu = 10\right)
\end{aligned}
$$

$$
\begin{aligned}
&= P\left(Z < \frac{\sqrt{n}\left(\mu_0 - \mu_A\right)}{\sigma} - 1.96 \mid \mu = 10\right) + P\left(Z > \frac{\sqrt{n}\left(\mu_0 - \mu_A\right)}{\sigma} + 1.96 \mid \mu = 10\right) \\
&= P\left(Z < -0.2933 \mid \mu = 10\right) + P\left(Z > 3.6267 \mid \mu = 10\right) \\
&= 0.386 + 0 \\
&= 0.386
\end{aligned}
$$