Hypothesis Testing

- The basic ingredients of a hypothesis test are
  1. the *null hypothesis*, denoted as H_o
  2. the *alternative hypothesis*, denoted as H_a
  3. the *test statistic*
  4. the *data*
  5. the *conclusion*.

- The hypotheses are usually statements about the values of one or more unknown parameters, denoted as θ here.
- The null hypothesis is usually a more restrictive statement than the alternative hypothesis, *e.g.* $H_o: \ \theta = \theta_o$, $H_a: \theta \neq \theta_o$.
- The burden of proof is on the alternative hypothesis.
- We will continue to believe in the null hypothesis unless there is very strong evidence in the data to refute it.
- The test statistic measures agreement of the data with the null hypothesis.
  - It is a reasonable combination of the data and the hypothesized value of the parameter.
  - It gets bigger when the data agrees less with the null hypothesis.

- When $\hat{\theta}$ is an estimator for θ with standard error $s_{\hat{\theta}}$, a common test statistic has the form

$$z = \frac{\hat{\theta} - \theta_o}{s_{\hat{\theta}}}.$$

- When the data agrees perfectly with the null hypothesis, $z = 0$.
- When the estimated and hypothesized values for θ become farther apart, z increases in magnitude.

There are two closely related approaches to testing.

1. One weighs the evidence against H_o.
2. The other ends in a decision to reject or not to reject H_o.

Weighing the evidence

- This approach uses the significance probability or P-value,
  - the probability of obtaining a value of the test statistic as or more extreme than the value actually observed, assuming that H_o is true.
  - This requires knowledge of the distribution of the test statistic under the assumption that H_o is true, the *null distribution* .

- For the two-sided alternative and test statistic mentioned above, the P-value is

$$P = 2Pr(z \geq |z_{observed}|)$$

- The factor 2 is required because *a priori* the sign of $z_{observed}$ is not known, and large (in magnitude) negative and positive values of z give evidence against H_o.
- Sometimes we use a one-sided alternative, $H_a : \theta > \theta_o$ or $H_a : \theta < \theta_o$.
- In these cases

$$P = Pr(z \geq z_{observed})$$

and

$$P = Pr(z \leq z_{observed})$$

respectively.

- The strength of the evidence against H_o is determined by the size of the P-value.
    - A smaller value for P gives stronger evidence.
- The logic is that if H_o is true, extreme values for the test statistic are unlikely, and therefore a possible indication that H_o is not true.
- By convention we draw the following conclusions

P value	Strength of evidence against H_o
$> .10$	none
$(.05, .10]$	weak
$(.01, .05]$	strong
$< .01$	very strong

- When $P < .01$, for example, we could say that ‘the results are statistically significant at the .01 level’, or ‘we have very strong evidence against the null hypothesis’.

Decision approach I

- The second approach to hypothesis testing requires a decision be made whether or not to reject H_o.
- One way to do this is to compare the P value to a small cut-off called the significance level α and to reject H_o if $P \leq \alpha$.
- Another way is to choose a *rejection region* and to reject H_o if the test statistic falls in this region.
- Two types of error are possible with this approach:
  1. A *type I error* occurs if H_o is rejected when it is true.
  2. A *type II error* occurs if H_o is not rejected when it is false.
- The type I error is considered to be much more important than the type II error.

Decision approach II

- A common analogy is with a court of law. In murder cases the presumption of innocence (H_o) is rejected only when the jury is convinced "beyond a shadow of a doubt" by very strong evidence (an extreme value for the test statistic).

- The type I error would be to convict and hang the accused (reject H_o) when he is innocent (H_o is true).
- The type II error, considered less serious, would be to let a guilty man go free (don't reject H_o when it is false).
- Recognizing the seriousness of the type I error, the rejection region is chosen so that the probability of rejecting H_o when it is true is a small value α.
- For example, the test statistic z discussed above frequently has an approximate normal distribution. For the two-sided alternative, with $\alpha = .05$, the rejection region consists of the values $|z| \geq z_{\alpha/2} = 1.96$.

Decision approach III

- When the data is assumed to be normally distributed and the variance is unknown and estimated by a sample variance, we use the t distribution.
- Finally, the data is collected and the test statistic is computed.
- If the test statistic falls in the rejection region we *reject* H_o *at level* α.
- Otherwise we *do not reject* H_o at level α.

- Remember that
  - A rejected H_o may in fact be true.
  - An H_o which is not rejected is probably not true either (This is why I *never* say 'H_o is accepted').
  - A result which is statistically significant (*i.e.* we have rejected H_o) may have no practical significance. With a very large sample size almost any H_o will be rejected.

Hypothesis testing in simple linear regression I

- The most common and useful test is whether or not the relationship between the response and predictor is significant.
- $H_0 : \beta_1 = 0$, *there is no linear relationship*
- $H_a : \beta_1 \neq 0$, *there is a linear relationship*
- The alternative is usually two sided.
- The test statistic is
$$T = \frac{\hat{\beta}_1}{se(\hat{\beta}_1)}$$
and this is compared to the t_{n-2} distribution.
- Here $se(\hat{\beta}_1) = s/\sqrt{S_{XX}}$.
- On occasion, we specify a value $\beta_{1,0}$ other than 0 in the null hypothesis.

Hypothesis testing in simple linear regression II

- Then the test statistic becomes

$$T = \frac{\hat{\beta}_1 - \beta_{1,0}}{se(\hat{\beta}_1)}.$$

- One can also test hypotheses about the intercept
- $H_0 : \beta_0 = \beta_{0,0}$,
- $H_a : \beta_0 \neq \beta_{0,0}$.
- Often we are interested in whether the intercept is zero, so $\beta_{0,0} = 0$.
- The test statistic is

$$T = \frac{\hat{\beta}_0 - \beta_{0,0}}{se(\hat{\beta}_0)}$$

and this is compared to the t_{n-2} distribution.

Tests on the mean of Y at x_0

- $H_0 : \mu_{x_0} = \mu_{x_0,0}$,
- $H_a : \mu_{x_0} \neq \mu_{x_0,0}$
- The test statistic is

$$T = \frac{\hat{\mu}_{x_0} - \mu_{x_0,0}}{se(\hat{\mu}_{x_0})}$$

and this is compared to the t_{n-2} distribution.