10.7 Comments on Hypothesis Testing

Note: When the test statistic does not fall in the RR, we do not say that we ________ H_0.

- Rather, we ________ H_0.
- When we ________, say, $H_0: \theta = \theta_0$, it does not imply that we are concluding that the true value of θ is θ_0.
- In fact, θ might not equal θ_0, but our test does not detect this difference (possibly because of a ________ or because the difference between θ and θ_0 is ________).

Note: When a hypothesis test rejects H_0, we say the result is ________.

- We should not confuse ________ with ________.

- Suppose we test $H_0: \mu_1 = \mu_2$ vs. $H_a: \mu_1 \neq \mu_2$, comparing mean reaction times between two populations, and we have very large samples.
- We might get a p-value of 0.02, implying ______________ significance if $\alpha = 0.05$.
- But if the sample mean reaction times only differ by 0.2 seconds, this difference may not be ______________ significant.
- A CI for $\mu_1 - \mu_2$ can better assess the practical size of the difference in this case.

10.8 Small-sample Tests for μ and for $\mu_1 - \mu_2$

- Section 10.3 covered tests for μ and $\mu_1 - \mu_2$ when the sample size(s) were large.
- Now suppose we have $Y_1, Y_2, \ldots, Y_n \overset{iid}{\sim} N(\mu, \sigma^2)$ where μ and σ^2 are unknown and n may be small.
- Consider testing $H_0: \mu = \mu_0$ against one of the three typical alternatives.

- If $H_0: \mu = \mu_0$ is true, then the statistic

has a ______________ with ______________ ______________ (shown in STAT 512).

- The test procedures in this situation are very similar to the z-tests from Sec. 10.3, but now we use:

Small-sample Test about μ

H_0	H_a	Test Statistic	RR

- Example 1: If the mean breaking strength for a manufactured ribbon is less than 185 pounds, it is too weak to sell.

An inspector selects 12 ribbons and tests their breaking strengths, finding $\overline{Y} = 180.4$ and $S = 8.21$ for this sample. Test using $\alpha = .05$.

Small-sample Test about $\mu_1 - \mu_2$

- Recall that if we have two independent samples $Y_{11}, Y_{12}, \ldots, Y_{1n_1} \overset{iid}{\sim} N(\mu_1, \sigma^2)$ and $Y_{12}, Y_{22}, \ldots, Y_{2n_2} \overset{iid}{\sim} N(\mu_2, \sigma^2)$, and if

then

- We can then test $H_0: \mu_1 - \mu_2 = \Delta_0$ with a <u>two-sample</u> <u>t-test</u>:

<u>Example 2:</u> A professor taught her two calculus classes using two different teaching methods (one traditional, one modern). She wanted to test whether the two methods produced significantly different mean final exam scores. The "traditional" class yielded $\bar{Y}_1 = 78.63$ and $s_1^2 = 24.62$ (with $n_1 = 12$) and the "modern" class ($n_2 = 18$) yielded $\bar{Y}_2 = 80.95$, $s_2^2 = 23.91$. Test using $\alpha = .05$.

- The t-tests are <u>robust</u>: If the assumptions behind them are moderately wrong, they still work approximately well:

- <u>If</u> the sample(s) are <u>moderately</u> non-normal, the t-tests can be used safely.

- If $\sigma_1 \neq \sigma_2$, the two-sample t-test can be used safely when $n_1 \approx n_2$.

10.9 Tests About Variances

- We now consider tests related to population variances.
- Let $Y_1, Y_2, \ldots, Y_n \overset{iid}{\sim} N(\mu, \sigma^2)$ with μ and σ^2 unknown.
- We test $H_0: \sigma^2 = \sigma_0^2$ against either an upper-tail, lower-tail, or two-tailed alternative.
- Recall that if $H_0: \sigma^2 = \sigma_0^2$ is true, then

- Let us use as our test statistic.

- Suppose we have $H_a: \sigma^2 > \sigma_0^2$.
- Then ________ values of S^2 favor H_a.

$\Rightarrow$ Reject H_0 when S^2 is ________.

$\Rightarrow$ Reject H_0 when

- Then the rejection region that keeps
$P[\text{Reject } H_0 \mid H_0 \text{ true}] = \alpha$ is:

Picture:

Similarly, for $H_a: \sigma^2 < \sigma_0^2$,

For $H_a: \sigma^2 \neq \sigma_0^2$,

Picture:

<u>Example 1:</u> A random sample of 20 sheet metal width measurements yielded a sample variance of 0.64. (Assume the data are normally distributed.) Test $H_0: \sigma^2 = 0.6$ vs. $H_a: \sigma^2 \neq 0.6$ using $\alpha = 0.05$.

Test to Compare Two Variances

- If we have two independent samples (of sizes n_1 and n_2) from $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$ populations, we can formally compare σ_1^2 and σ_2^2.

- Recall that

- So dividing by the respective d.f. and taking the ratio, we see:

has an ________________ with ________________

and ____________________________.

- If $H_0: \sigma_1^2 = \sigma_2^2$ is true, then

- So can be our test statistic.

- If $H_a: \sigma_1^2 > \sigma_2^2$, then ________ values of F favor H_a.

- RR :

- If H_a is in the reverse direction, simply reverse the labeling of populations 1 and 2 and conduct the upper-tail test.

- If $H_a: \sigma_1^2 \neq \sigma_2^2$, then ____________ values of F favor H_a.

- Note that both F and $\frac{1}{F}$ have F-distributions, but with the d.f. reversed.

- So we can reject H_0 if

RR:

Example 2: Suppose two (independent) normal samples of 11 and 13 measurements yielded $s_1 = 8.62$ and $s_2 = 9.13$. Test $H_0: \sigma_1^2 = \sigma_2^2$ vs. $H_a: \sigma_1^2 \neq \sigma_2^2$ using $\alpha = .05$.