MATH 243, LECTURE 20

1. Comparing (means of) two samples

We continue adding levels of detail so our procedures may apply in a larger number of real-world settings. As we have said many times, the best data often come from experiments, where there are two populations being treated by different means, and then measured and compared. Before we used single-variable statistics to analyze a matched-pair experiment. But matched-pair experiments have both practical and theoretical drawbacks, so we develop statistics which address more general experimental design.

As before we should think about means and deviations, but the formulae can now be complicated.

What we want to understand: the difference of means, $\mu_1 - \mu_2$. The basic measurement we start with: the difference of observed means $\bar{x_1} - \bar{x_2}$.

In order to find confidence intervals and test hypotheses, we need to understand standard deviations and errors. Fortunately some sharp mathematicians and statisticians come to our rescue.

Theorem 1. *If two distributions of size n_1 and n_2 have standard deviations σ_1 and σ_2, then the deviation for the observed difference is $\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$.*

We get the *standard error*, called SE, for two samples by using s_1 and s_2 in place of σ_1 and σ_2.

We may use a t-distribution to approximate statistics which approximate the difference of means $\mu_1 - \mu_2$. Ignoring this distinction between one- and two- variable distributions (which are taken into account by sophisticated statistics programs) we have the following.

Theorem 2. *The following approximations may be used when t-procedures are applicable.*

- *With probability $C\%$ the (true) difference of means $\mu_1 - \mu_2$ has values between*

$$(\overline{x_1} - \overline{x_2}) - t^* SE \text{ and } (\overline{x_1} - \overline{x_2}) + t^* SE,$$

  where t^* is the critical value associated to the $t(n-1)$-distribution where n is the smallest of n_1 and n_2.
- *To test the null hypothesis, $H_0 : \mu_1 = \mu_2$, we calculate*

$$t_0 = \frac{\overline{x_1} - \overline{x_2}}{SE}$$

  Then $P(t \geq |t_0|) < \alpha$ supports the alternative hypothesis $H_a : \mu_1 \neq \mu_2$ at level α.

To summarize, if we use the standard error SE in the places the single-sample standard error was used, we may use the same methods to understand the true difference of means from the observed difference of means. (It's the same old song...)

1.1. Examples of hypothesis testing for two means.

Example 3. *Estimate $\mu_1 - \mu_2$ giving a confidence interval of level 95% when we have:*

- *A sample of size 19 from population A, with mean 54, and sample standard deviation 5.*
- *A sample of size 23 from population B, with mean 49 and sample standard deviation 4.*

Next we give all details in doing Exercise 17.38 from the text. This exercise gives some IQ data for some boys and girls from the same midwestern school district and asks if there is a statistically significant difference between the means. After keying some numbers into a calculator, we get the following information for our two samples:

Population	Mean	Sample Size	Sample mean	Sample s.d.
Girls	μ_1	31	$\overline{x}_1 = 105.84$	$s_1 = 14.27$
Boys	μ_2	47	$\overline{x}_2 = 110.96$	$s_2 = 12.12$

(1) Our null hypothesis is that boys' IQ scores are the same as girls' IQ scores. That is

$$H_0 : \mu_1 = \mu_2.$$

Our alternative hypothesis is that boys have higher IQ scores.

$$H_a : \mu_1 < \mu_2 \text{ or } \mu_1 - \mu_2 < 0.$$

We wish to test this using our data.

(2) We calculate our two-sample t-statistic

$$t = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{105.84 - 110.96}{\sqrt{6.569 + 3.125}} = \frac{-5.12}{3.114} = -1.644.$$

(3) We calculate our P-value. Since H_a is $\mu_1 - \mu_2 < 0$, we wish to look for $P(t \leq -1.644)$. We use $t(30)$ since 31 is our smaller sample size. (see p. 452 for a more accurate way to determine degrees of freedom).

From the calculator, $P(t \leq -1.644) = .0553$.

(4) We draw our conclusion: If we assume H_0 is *true*, then the probability of seeing samples like the ones we have is .0553. This is moderately low, so our assumption that H_0 was true is probably wrong. So, this is moderate evidence that boys score higher on IQ tests than girls. Which is in turn evidence that small differences in tests such as IQ tests do not accurately reflect much of anything.

(5) We can ask the calculator to *do* the test for us. This is under $\boxed{\text{STAT}}$, TESTS, 4:2-SampTTest. We get $df = 56.93$, $t = -1.64$, $P = .053$.

We *still* need to do step 4 (conclusion) above. And we need to do it carefully, because we've possibly lost track of what all our numbers mean.

Example 4. *Physicians wish to measure the effectiveness of leech therapy on arthritis pain (I'm not making this up! See*
`http://www.annals.org/cgi/content/full/139/9/724`*)*
They took a sample of 51 patients. They assigned a group of 24 patients to receive "leech therapy," and a control group of 27 patients to receive a conventional pain-relief therapy (diclofenac gel).

One week after the leech treatment (it was one treatment lasting a little over an hour involving 4 to 6 leeches), the leech group had a mean pain index of 19.3 with a standard deviation of 12.2.

After one week of the other treatment*, the control group had a mean pain index of 42.4 with standard deviation of 19.7.*

Test the null hypothesis that the effect of treatment by leeches is the same as the effect of conventional treatments.