

Stat 13, Tue 5/22/12.

1. Midterms back.
2. SE and CI for difference between two means.
3. Testing basics.
4. Z test and t test for two samples.

Final exam is Thur, 6/7, in class. Hw6 is due this Thu, 5/24.

Change in hw6: Ignore problem 7.30 in the 3<sup>rd</sup> edition / 7.2.8 in the 4<sup>th</sup> edition.

Be SILENT while I am passing out the exams or I will ask you to leave.

Gradegrubbing procedure, again: if you would like a question (or more than one question) reevaluated, submit your exam or homework and a **WRITTEN** explanation of why you think you deserve more points and how many more points you think you deserve **to your TA**. The TA will then give it to me, and I will consider it, and then give it back to the TA to give back to you.

Midterms: 90-100 = A range, 80-90 = B range, 70-80 = C range, 60-70 = D range, < 60 = F.

I record your number, not the letter grade. Your grade is on the last page.

Mean = 83. SD = 19.

Common things:

- \* 2. 95% of the values for the normal distribution are in  $\mu \pm 1.96 \sigma$ .  $1.96(5.2) \sim 10.2$ .
- \* 8. I messed up here and said rabbit lengths instead of rabbit hair counts.
- \* 14. Let  $X_1 = 1$  if the 1<sup>st</sup> sample mean is < 102.4, and 0 otherwise. Similarly for  $X_2$  and  $X_3$ .  
 $X = X_1 + X_2 + X_3$ .  $E(X) = E(X_1) + E(X_2) + E(X_3) = 88.49\% + 88.49\% + 88.49\% \sim 2.655$ .
- \* Do not guess NOTA. I never, ever intend for the correct answer to be NOTA. So, if you're getting an answer that is not there, check to see if you've made a mistake.

## 2. SE and CI for difference between two means.

Suppose you have two samples: a SRS of size  $n_1$  from population 1 and a SRS of size  $n_2$  from population 2.

For instance, pop. 1 is those drinking 4-6 cups of coffee daily, and pop. 2 is those drinking 0 cups a day.

For each person, record blood pressure (BP).

Let  $\mu_1$  be the mean BP of population 1, and  $\mu_2$  is the mean BP of population 2.

Let  $\bar{x}_1$  be the mean BP of sample 1, and  $\bar{x}_2$  is the mean BP of sample 2.

Often we're interested in estimating the difference,  $\mu_1 - \mu_2$ .

Our estimate would be  $\bar{x}_1 - \bar{x}_2$ . But what if we want a 95% CI for  $\mu_1 - \mu_2$ ?

We'd simply use be  $\bar{x}_1 - \bar{x}_2 \pm 1.96 \text{SE}_{\text{diff}}$ ,

where  $\text{SE}_{\text{diff}}$  is the standard error for the difference, i.e. the typical distance between the difference between sample means and the difference between the two pop.means.

There's an easy formula.  $\text{SE}_{\text{diff}} = \sqrt{(\text{SE}_1^2 + \text{SE}_2^2)}$ ,

where  $\text{SE}_1 = \frac{\sigma_1}{\sqrt{n_1}}$  is the SE for sample mean 1 and  $\text{SE}_2 = \frac{\sigma_2}{\sqrt{n_2}}$  is the SE for sample 2.

So  $\text{SE}_{\text{diff}} = \sqrt{(\text{SE}_1^2 + \text{SE}_2^2)} = \sqrt{(\sigma_1^2/n_1 + \sigma_2^2/n_2)}$ .

Usual CLT assumptions must hold for both samples, and if  $n$  is large, plug in  $s$  for  $\sigma$ .

For example, suppose sample 1 is an SRS of 100 coffee drinkers, and in this sample,  $\bar{x}_1 = 124$  and  $s_1 = 10$ . Sample 2 is an SRS of 85 non-drinkers, and  $\bar{x}_2 = 127$ ,  $s_2 = 8$ .

Then a 95% CI for  $\mu_1 - \mu_2$  would be  $\bar{x}_1 - \bar{x}_2 \pm 1.96 \text{SE}_{\text{diff}}$ ,

which is  $124 - 127 \pm 1.96 \sqrt{(10^2/100 + 8^2/85)} = -3 \pm 2.60$ . Note: it doesn't cover 0.

If both populations are known to be normally distributed, and  $n_1$  and  $n_2$  are small, and they're both SRSs, then you'd substitute  $z^* = 1.96$  (for a 95% CI) for  $t^*$ , where now the degrees of freedom,  $df = n_1 + n_2 - 2$ .

Ignore Formula 6.7.1 on p227 in the 3<sup>rd</sup> edition = Formula 7.1 in the 4<sup>th</sup> edition, involving the df for the difference between two means. Use  $n_1 + n_2 - 2$  instead.

A 95% CI for  $\mu_1 - \mu_2$  would be  $\bar{x}_1 - \bar{x}_2 \pm t^* SE_{diff}$ .

For instance, with all the numbers the same as in the last example, but if both populations were known to be normally distributed, and  $n_1 = 10$  and  $n_2 = 7$ ,  $df = 17 - 2 = 15$ , so  $t^* = 2.131$ , and now  $SE_{diff} = \sqrt{(10^2/10 + 8^2/7)} = 4.375$ , and a 95% CI for  $\mu_1 - \mu_2$  would be  $124 - 127 \pm 2.131(4.375) = -3 \pm 9.323$ .

The exact same formula holds when both samples are of 0-1 data.

For instance, say you record 1 if the person dies and 0 if he/she survives the 15 year period. In the SRS of 100 coffee drinkers, 18% die, and in the SRS of 85 nondrinkers, 23% die.

A 95% CI for the difference between the two population percentages is...

$$\begin{aligned} & 18\% - 23\% \pm 1.96 \sqrt{(s_1^2/n_1 + s_2^2/n_2)}, \quad [s_1 = \sqrt{(18\% \times 82\%)}, \text{ so } s_1^2 \text{ simply} = 18\% \times 82\%] \\ & = -5\% \pm 1.96 \sqrt{(18\% \times 82\%/100 + 23\% \times 77\%/85)} \\ & = -5\% \pm 11.69\%. \end{aligned}$$

So, this time it contains 0, so we can't rule out the possibility that there is no difference between the two population percentages. That is, the difference between the two sample percentages might be due to chance alone.

### 3. Testing basics.

Suppose you take two SRSs from two populations and you want to test whether the population means might really be the same.

e.g. one sample mean = 18% and the other is 23%. You want to test whether something like this could reasonably have happened just by chance alone, if the populations were actually identical. Otherwise we conclude that the two population means are probably not equal.

There are different tests, but we'll just talk about the Z-test (or normal test) and t test.

Assumptions: same as before. For each sample, it must be

SRS (or obs are known to be independent)

AND  $n$  is large (or pop is known to be normally distributed).

If in addition, for at least one of the samples,  $n$  is small, pop. is normal, and  $\sigma$  is unknown, then use  $t$  instead of  $Z$ .

After checking assumptions, the remaining steps in testing are

- \* stating the hypotheses,
- \* computing the test statistic ( $Z$  or  $t$ ),
- \* computing the p-value, and
- \* concluding.

### Hypotheses.

Let  $\mu_1$  and  $\mu_2$  be the means of the populations from which the samples are drawn.

Null hypothesis ( $H_0$ ):  $\mu_1 = \mu_2$ .

This means that the observed difference between  $\bar{x}_1$  and  $\bar{x}_2$  is due to chance alone.

Alternative hypothesis ( $H_a$ ):  $\mu_1 \neq \mu_2$ . Difference is not due to chance alone. (2-sided test.)

Or  $H_a: \mu_1 > \mu_2$ . Or  $H_a: \mu_1 < \mu_2$ . (1-sided tests). Direction depends on your data.

When in doubt, do a two-sided test, unless I specifically say to do a 1-sided test.

Usually we specify these hypotheses numerically.

### Z-statistic.

A test statistic is a summary of the data.

Z-statistic =  $(\bar{x}_1 - \bar{x}_2) \div SE_{\text{diff}}$ . Same for t, but call it t! If either  $\sigma$  is unknown, just plug in s.

### P-value.

The p-value is the probability, assuming  $H_0$  is true, that the test statistic will be at least as extreme as that observed.

Use normal table, or with the t table, figure out if the t-statistic is bigger than the critical value  $t^*$  for your corresponding value of  $\alpha$ , which is the significance level of the test (typically  $\alpha$  is 5%). For a one-sided test, cut the p-value in half.

**Conclusion.** Small p-value = strong evidence against  $H_0$ . Large p-value = weak evidence against  $H_0$ . The observed difference is *statistically significant* if the p-value < 5%.

The observed difference is *highly significant* if the p-value < 1%.

Examples.

Again, suppose sample 1 is an SRS of 100 coffee drinkers, and in this sample,  $\bar{x}_1 = 124$  and  $s_1 = 10$ . Sample 2 is an SRS of 85 non-drinkers, and  $\bar{x}_2 = 127$ ,  $s_2 = 8$ . Do a test at significance level  $\alpha = 5\%$  to see if the two sample means are significantly different.

Check assumptions. Both SRSs and  $n$  is large in each.

$H_0: \mu_1 = \mu_2$ .

$H_a: \mu_1 \neq \mu_2$ . (two-sided). If one-sided, it'd be  $H_a: \mu_1 < \mu_2$ .

$$z = (\bar{x}_1 - \bar{x}_2) \div SE_{\text{diff}} = (124 - 127) \div \sqrt{(10^2/100 + 8^2/85)} = -3 \div 1.324 = -2.27.$$

So, the p-value for this 2-sided test is

$$P(\text{std normal is} < -2.27) + P(\text{std normal is} > 2.27) = 2 \times 0.0116 = .0232 = 2.32\%.$$

The p-value is smaller than 5%, so the difference is statistically significant. This means the evidence against the null hypothesis is strong, i.e. we have strong evidence suggesting that coffee drinkers and non coffee drinkers have different average BPs. The difference (between the sample means of 124 and 127) is statistically significant, and is very unlikely to be due to chance alone.

Note that sample size is a factor. Sometimes a difference of -3 is statistically significant, and sometimes it's not. It depends on the sample size and the SDs.

An example using t, and where the first sample mean is bigger, and where it's not stat. sig. Suppose sample 1 is an SRS of 10 coffee drinkers, and in this sample,  $\bar{x}_1 = 127$  and  $s_1 = 10$ . Sample 2 is an SRS of 8 non-drinkers, and  $\bar{x}_2 = 124$ ,  $s_2 = 7$ . Both populations are normal. Do a test at significance level  $\alpha = 5\%$  to see if the two sample means are significantly different.

Check assumptions. Both SRSs, pops are normal, n is small,  $\sigma$ s are unknown, so do a t test.

Ho:  $\mu_1 = \mu_2$ .

Ha:  $\mu_1 \neq \mu_2$ . (two-sided). If one-sided, we would write Ha:  $\mu_1 > \mu_2$ .

$t = (\bar{x}_1 - \bar{x}_2) \div SE_{\text{diff}} = (127 - 124) \div \sqrt{(10^2/10 + 7^2/8)} = 3 \div 4.016 = 0.747$ .

Using the t-table, we can't figure out the exact p-value for this test but we can find the critical value  $t^*$ .  $df = n_1 + n_2 - 2 = 16$ , and  $\alpha = 5\%$  so  $t^*_{16} = 2.120$ .

Our observed  $t = 0.747$ .

If our  $t = 2.120$ , then our p-value would be 5%.

If our  $t > 2.120$ , then our p-value would be  $< 5\%$ .

Here, our  $t < 2.120$ , so our p-value is  $> 5\%$ . This means that the observed difference is not statistically significant. The evidence is weak against the null hypothesis that coffee drinkers and non coffee drinkers have the same average BPs. The difference (between the sample means of 127 and 124) is not statistically significant, and could plausibly be attributed to chance alone.