Real One- and Two-Sample Statistical Inference

This session is "a version of MMD&S Ch 7 plus some." It

1. fixes the "known σ" requirement of the one-sample methods of the previous session/Ch 6
2. notes the application of one-sample methods to "paired data" contexts
3. introduces two-sample inference methods of two varieties
   - the MMD&S "preferred"/default method
   - the "pooled s" method

One Sample Methods Based on s rather than σ

Confidence Intervals for μ

The introduction to probability-based inference in the previous session was based mostly on the fact that the sampling distribution of

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

is (at least approximately) standard normal. The σ in that formula propagates through the inference formulas and makes them mathematically OK, but of limited practical value. It would be nice if one could begin in a way that σ didn't get involved. In fact, it would be nice if one could replace σ with s above and still use basically the same logic as before. Happily this can be done.

There is the following probability fact:

> When sampling from a normal population/universe/process, the random quantity
>
> $$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$
>
> has a famous standard probability distribution called the "t distribution with $df = n - 1$"

The t distributions are tabled on the inside back cover of MMD&S and are pictured in Figure 7.1, page 434. These are bell-shaped, centered at 0, and "flatter than" the standard normal distribution. For large values of df (degrees of freedom) they are virtually indistinguishable from the standard normal distribution.

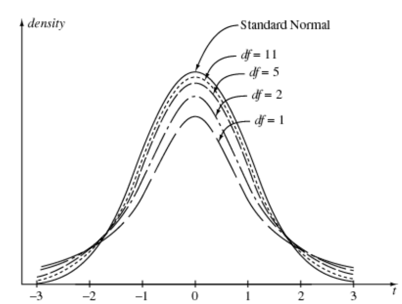

Figure 1: Four t Distributions and the Standard Normal Distribution

Tables for t distributions can not be as complete as ones for the standard normal distribution, as there is a different t distribution for each value of df. The MMD&S table is set up so that one looks for a right-tail area on the top margin of the table and the appropriate value for df on the left margin, and then reads the associated cut-off value from the body of the table. This is pictured below in schematic fashion.

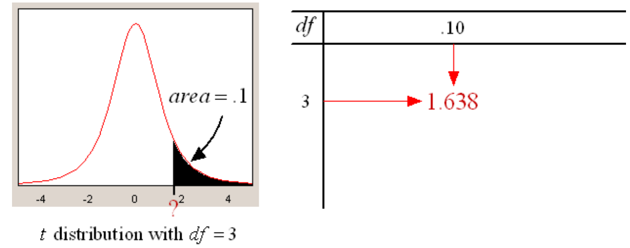

Figure 2: Use of the t Table in MMD&S

Example A data set in Dielman's *Applied Regression Analysis* (taken originally from *Kiplinger's Personal Finance*) gives rates of return for $n = 27$ no-load mutual funds in 1999. Suppose that

1. rates of return for such funds in 1999 were approximately normally distributed
2. the Dielman data set is based on a random sample of no-load mutual funds

Consider inference for the mean rate of return, μ, for such funds in 1999. From the t table

$$P\left(-2.056 < (\text{a } t \text{ random variable with } df = 26) < 2.056\right) = .95$$

(Look in the table under a right-tail area of .025.)

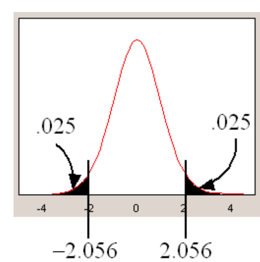

Figure 3: t Distribution With $df = 26$

Then for 95% of all samples of $n = 27$ no-load funds,

$$-2.056 < \frac{\bar{x} - \mu}{\frac{s}{\sqrt{27}}} < 2.056$$

But this is algebraically the same as

$$\bar{x} - 2.056\frac{s}{\sqrt{27}} < \mu < \bar{x} + 2.056\frac{s}{\sqrt{27}}$$

So, for samples of size $n = 27$, 95% confidence limits for μ are

$$\bar{x} \pm 2.056\frac{s}{\sqrt{27}}$$

The general formula for confidence limits for μ (based on a sample of size n from a normal universe) is

$$\bar{x} \pm t\frac{s}{\sqrt{n}}$$

where t is a percentage point from the t distribution with $df = n - 1$.

Example For the no-load mutual funds, the Dielman data set has $n = 27, \bar{x} = 13.3$ and $s = 4.1$. So 95% confidence limits for the mean rate of return are

$$13.3 \pm 2.056\frac{4.1}{\sqrt{27}}$$

that is,

$$13.3 \pm 1.6$$

Demonstration As an example, we will make some 80% confidence interval estimates of the mean of the brown bag based on samples of size $n = 5$ (without making use of the population standard deviation). The limits will be

$$\bar{x} \pm t\frac{s}{\sqrt{n}}$$

which in this case is

$$\bar{x} \pm 1.533\frac{s}{\sqrt{5}}$$

which is

$$\bar{x} \pm .69s$$

Sample	Values	$\bar{x}$	s	$.69s$	Successful?
1					
2					
3					
4					
5					

Note that

- the 80% figure is a lifetime batting
- the length of the interval is random, changing sample to sample

MMD&S call

$$\frac{s}{\sqrt{n}}$$

"the standard error of the mean." "Standard error" is synonymous with "estimated standard deviation." $s/\sqrt{n}$ is an estimated version of the theoretical standard deviation of $\bar{x}$,

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Exercise Problem 7.31 page 456 MMD&S

Additional comments on the "t interval for μ":

- The text talks about this method being "robust." That is, it doesn't go terribly wrong unless the population being sampled is "radically non-bell-shaped." Strictly speaking, the "95% confidence" guarantee refers to applications where the population being sampled is normal ... but "robustness" says that it isn't ridiculously wrong unless the population is badly non-normal. (What I think is a 95% interval might in reality only be a 90% interval, but it won't be a 15% interval.)

- For large n, the formula turns into
$$\bar{x} \pm z\frac{s}{\sqrt{n}}$$
and some authors call this the "large sample confidence interval for μ."

Prediction Intervals for x_{new}

One can "fix the known σ problem" for the prediction intervals of Session 3 ... but only for for cases where one is reasonably sure that the universe being sampled is fairly bell-shaped. That is (both the "known σ" method of Session 3 and) what follows here is *not* robust. I would only apply this material in practice if I somehow "know" the *population* is approximately normal ... probably by checking the shape of a histogram for the *sample*.

The basis of making prediction intervals is the probability fact:

> When sampling from a normal population/universe/process, the random quantity
> $$t = \frac{x_{\text{new}} - \bar{x}}{s\sqrt{1+\frac{1}{n}}}$$

has the t distribution with $df = n - 1$

This leads to prediction limits for x_{new}
$$\bar{x} \pm ts\sqrt{1+\frac{1}{n}}$$

Example For the Dielman no-load mutual funds (1999 rates of return), consider a 95% prediction interval for the rate of return for a single additional fund not included in the original data. Recall that
$$n = 27, \bar{x} = 13.1, \text{ and } s = 4.1$$
Assuming that the population is reasonably bell-shaped, we would therefore use limits
$$13.1 \pm 2.056\,(4.1)\sqrt{1+\frac{1}{27}}$$

since the upper 2.5% point of the t distribution with $df = 26$ is 2.056. This is
$$13.1 \pm 8.6$$

NOTICE that although this interval is centered at 13.1 just like the confidence interval for μ, it is much wider (the earlier one had limits 13.3 ± 1.6). This interval is doing a different job and MUST be wider ... it is not attempting to bracket μ, it is attempting to bracket x_{new}. Even if one knew μ and σ exactly (which one doesn't) there would still be uncertainty and the need for a wide interval in prediction.

Demonstration As an example, we will make some 80% prediction intervals for sampling from the brown bag based on samples of size $n = 5$ (without making use of the population standard deviation). The limits will be
$$\bar{x} \pm ts\sqrt{1+\frac{1}{n}}$$

which in this case is

$$\bar{x} \pm 1.533 s \sqrt{1 + \frac{1}{5}}$$

i.e.

$$\bar{x} \pm 1.68 s$$

Sample	Values	$\bar{x}$	Endpoints	x_{new}	Successful?
1					
2					
3					
4					
5					

The 80% guarantee is a lifetime batting average guarantee *for the whole business of selecting 5, making an interval, selecting one more, and seeing if the*

new one is in the interval.

Be warned that prediction interval methods are *not* robust. The relevance of the nominal confidence guarantee depends critically on the normality of the distribution being sampled.

Exercise For the scenario of Problem 7.31 on page 456 of MMD&S, make 95% and 99% prediction intervals for one more corn price.

Significance Testing for μ

Significance testing for μ without the "known σ" assumption goes just as you should expect. You replace σ (in the Session 3 method) with s, and z with t,

and operate as before. To test H_0:$\mu = \#$ use the test statistic

$$t = \frac{\bar{x} - \#}{\frac{s}{\sqrt{n}}}$$

and the t table (with $df = n - 1$) to get p-values.

Example For the Dielman 1999 no-load mutual fund problem, let's ask

"Was the mean rate of return in 1999 clearly below 15?"

This can be translated as an invitation to test the hypotheses

$$H_0{:}\mu = 15$$
$$H_a{:}\mu < 15$$

With $n = 27, \bar{x} = 13.3$ and $s = 4.1$ the observed value of the test statistic is

$$t = \frac{\bar{x} - \#}{\frac{s}{\sqrt{n}}} = \frac{13.3 - 15}{\frac{4.1}{\sqrt{27}}} = -2.15$$

which is to be compared to tabled values for the t distribution with $df = 26$. Pictorially this is

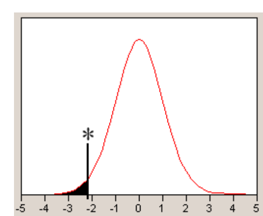

Figure 4: p-value for Testing $H_0 : \mu = 15$ vs $H_a : \mu < 15$

Since the upper 2.5% point of the t distribution with $df = 26$ is 2.056 and the upper 2% point is 2.162, the p-value is between .02 and .025. There is reasonably convincing evidence that in fact $\mu < 15$.

Exercise Suppose that for the scenario of Problem 7.31 on page 456 of MMD&S, if the mean price is clearly below \$2.10, policy makers want to intervene. **Assess the strength of the sample evidence that** the mean price is below \$2.10. (The phrase in bold is standard "code" indicating that what follows should be placed in the alternative hypothesis).

Application of One-Sample t Methods to Paired Data

An important application of the foregoing is to "paired data." "Paired data" arise where one has two similar measurements from a single sample of n objects/

items/subjects. These might be "before and after" or " left side and right side" or "with and without treatment" values on a single set of units. A schematic is

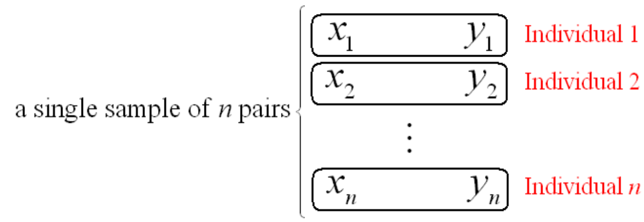

Figure 5: Paired Data

A standard method of analysis is to take differences

$$x - y = d$$

and to do inference for μ_d. See, for example, Problem 7.42 page 459 of MMD&S. The key to understanding when this method is appropriate is that it applies to a *single* sample of objects, each of which has two associated measurements. This is something much different from the material in Section 7.2 of MMD&S, that concerns comparisons made on the basis of *two samples* of objects.

Two-Sample Inference for $\mu_1 - \mu_2$

The problem treated in Section 7.2 of MMD&S is that of comparing two population/process/distribution means based on samples from those. Schematically,

one has

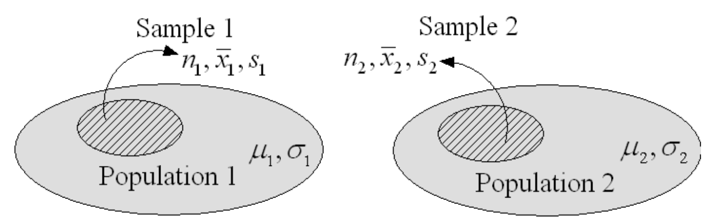

Figure 6: Comparison of Two Population Means on the Basis of Two Samples

and the object is to use the sample information

$$n_1, \bar{x}_1, s_1 \text{ and } n_2, \bar{x}_2, s_2$$

to compare the population means μ_1 and μ_2.

Text's Favorite/Default Method of Inference for $\mu_1 - \mu_2$

MMD&S's favorite method for making confidence intervals for and testing hypotheses about $\mu_1 - \mu_2$ is based on the variable

$$t = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

This is not really t-distributed (even if the populations sampled are normal). BUT, if one acts as if it were t-distributed with

$$df = \text{the smaller of } n_1 - 1 \text{ and } n_2 - 1$$

one gets conservative inferences (actual confidence levels at least the size of nominal ones and p-values not smaller than what really ought to be stated).

Treating this quantity as a t-variable leads to confidence limits for $\mu_1 - \mu_2$ of the form

$$\bar{x}_1 - \bar{x}_2 \pm t\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

for df =the smaller of $n_1 - 1$ and $n_2 - 1$.

Demonstration Make an 80% confidence interval for the difference in means of the blue and brown bags

$$\mu_{\text{blue}} - \mu_{\text{brown}}$$

based on a sample of size $n_{\text{blue}} = 5$ from the blue bag and a sample of size $n_{\text{brown}} = 4$ from the brown bag. (The bags actually have $\mu_{\text{blue}} = 10, \sigma_{\text{blue}} = 3.47, \mu_{\text{brown}} = 5$, and $\sigma_{\text{brown}} = 1.715$.)

	Sample Values	$\bar{x}$	s
Blue			
Brown			

Notice that the upper 10% point of the t distribution with

$$\begin{aligned} df &= \text{smaller of } (5-1) \text{ and } (4-1) \\ &= 3 \end{aligned}$$

is 1.638. So approximate/conservative 80% limits for $\mu_{\text{blue}} - \mu_{\text{brown}}$ are

$$(\quad - \quad) \pm 1.638\sqrt{\frac{\quad}{5} + \frac{\quad}{4}}$$

and in this made-up example, one can check to see whether these limits enclose the true difference $\mu_{\text{blue}} - \mu_{\text{brown}} = 10 - 5 = 5$. The 80% confidence level is a guaranteed lifetime batting for repeated applications of selecting two samples, making an interval, and checking to see whether it includes 5.

Exercise Do part c) of problem 7.79 page 485 of MMD&S. Note that here

$$\bar{x}_{\text{women}} = 141.1 \quad \bar{x}_{\text{men}} = 121.3$$
$$s_{\text{women}} = 26.4 \quad s_{\text{men}} = 32.9$$

The significance testing version of the MMD&S default method of inference for $\mu_1 - \mu_2$ is to test

$$\text{H}_0\text{:}\mu_1 - \mu_2 = \#$$

(# is often 0, indicating *no difference* in the two means, μ_1 and μ_2) using the test statistic

$$t = \frac{\bar{x}_1 - \bar{x}_2 - \#}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

and getting conservative p-values from the t table with df =the smaller of $n_1 - 1$ and $n_2 - 1$.

An Alternative Method of Inference for $\mu_1 - \mu_2$

A second method of inference is based on the assumption that the two distributions being sampled have the same variability, i.e. that

$$\sigma_1 = \sigma_2 = \sigma$$

In practice, this works fine as long as the two standard deviations are not radically different.

When two samples come from populations with the same standard deviation it makes sense to somehow "pool" the two sample standard deviations to make a single estimate of the supposedly common population standard deviation. A mathematically convenient form for such a pooled estimate is

$$s_{\text{pooled}} = \sqrt{\frac{(n_1 - 1)\, s_1^2 + (n_2 - 1)\, s_2^2}{(n_1 - 1) + (n_2 - 1)}}$$

Example A famous 1960's marketing study compared the mean ages of purchasers and non-purchasers of Crest toothpaste. Summary statistics for the

two samples of toothpaste buyers were

Non-Purchasers	Purchasers
$n_1 = 20$	$n_2 = 20$
$\bar{x}_1 = 47.2$	$\bar{x}_2 = 39.8$
$s_1 = 13.62$	$s_2 = 10.04$

The two sample standard deviations are not terribly different, so it is perhaps not unreasonable to assume that the variability in ages of non-purchasers is the same as that for purchasers. Then

$$\begin{aligned} s_{\text{pooled}} &= \sqrt{\frac{(20-1)\,(13.62)^2 + (20-1)\,(10.04)^2}{(20-1)+(20-1)}} \\ &= 11.96 \text{ years} \end{aligned}$$

is a pooled estimate of the standard deviation of age for either group.

Notice that

$$s_{\text{pooled}} = \sqrt{\text{a weighted average of } s_1^2 \text{ and } s_2^2}$$

and is always a number between s_1 and s_2.

Then (having prepared the notion of pooling sample standard deviations to produce the quantity s_{pooled}) the basic variable used in inference for $\mu_1 - \mu_2$ under an assumption that $\sigma_1 = \sigma_2$ is

$$t = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{s_{\text{pooled}}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

This variable has exactly a t distribution with

$$\begin{aligned} df &= (n_1 - 1) + (n_2 - 1) \\ &= n_1 + n_2 - 2 \end{aligned}$$

when sampling from two normal distributions with $\sigma_1 = \sigma_2$.

This basic probability fact leads to confidence limits for $\mu_1 - \mu_2$

$$\bar{x}_1 - \bar{x}_2 \pm t s_{\text{pooled}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

for t a percentage point of the t distribution with $df = n_1 + n_2 - 2$. Further, $\text{H}_0\text{:}\mu_1 - \mu_2 = \#$ can be tested using the test statistic

$$t = \frac{\bar{x}_1 - \bar{x}_2 - \#}{s_{\text{pooled}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

where p-values are taken from the t distribution with $df = n_1 + n_2 - 2$.

Example In the Crest marketing study, 95% confidence limits for the difference

in mean ages of non-purchasers and purchasers are

$$\bar{x}_{\text{non}} - \bar{x}_{\text{Crest}} \pm t s_{\text{pooled}} \sqrt{\frac{1}{n_{\text{non}}} + \frac{1}{n_{\text{Crest}}}}$$

that is

$$47.2 - 39.8 \pm 2.024\,(11.96) \sqrt{\frac{1}{20} + \frac{1}{20}}$$

i.e.

$$7.4 \pm 7.66$$

(The 2.024 value is the upper 2.5% point of the t distribution with $df = 38$ (obtained from a statistical program since the MMD&S table doesn't list that number of degrees of freedom).)

The Crest study doesn't definitively establish a difference in mean ages for the groups. The uncertainty/margin of error (7.66) exceeds the magnitude of

the observed difference in $\bar{x}$'s (7.4). That is, the confidence interval for the difference includes not only positive values, but 0, and some negative values as well. A significance test of $\text{H}_0\text{:}\mu_{\text{non}} - \mu_{\text{Crest}} = 0$ versus $\text{H}_a\text{:}\mu_{\text{non}} - \mu_{\text{Crest}} \neq 0$ would produce a p-value larger than .05. (The 95% confidence interval for $\mu_{\text{non}} - \mu_{\text{Crest}}$ includes 0.)