One and Two-Sample T-tests

Lectures 16, 17, & 18

Readings: GW 9 & 10 & SS 9 & 10

Single Sample Inference when True Variance is not know (almost always)

- Most typically, since we don't know μ, we don't know σ either.
- What do we know from our sample data that is an unbiased estimate of σ ?
- Sample standard deviation, s!
- Introduces additional uncertainty into test. Need to compensate by using a different statistic with more extreme critical values, to the extent that additional uncertainty is introduced.
- When will values of s be closer on average to the values of σ ?

Single Sample Inference when True Variance is not know (almost always)

- Estimated standard error of the mean

$$s_{\bar{X}} = \frac{s}{\sqrt{n}} \approx \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

- Approximation depends on sample size
  - (actually df = n - 1)
- Computation of test statistic unchanged, but call it a t statistic with df = (n - 1)

$$t_{n-1} = \frac{\bar{X}-\mu}{s_{\bar{X}}} = \frac{\bar{X}-\mu}{s/\sqrt{n}}$$

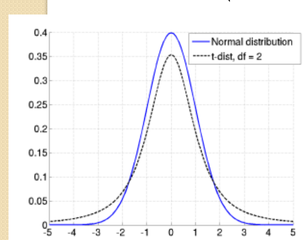
- Critical value of Z was ±1.96. Critical value of t depends on df.

Distribution of t-statistic

$$t = \frac{M-\mu_0}{\sqrt{\frac{s^2}{n}}}$$

magnitude

standard err



- This distribution is the t-distribution, and its shape depends on the degrees of freedom, n - 1.
- Write this as t(df) or t_{df}

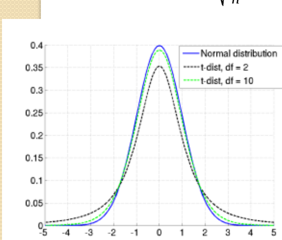
t(20) = 1.5
or
t_{20} = 1.5

Distribution of t-statistic

$$t = \frac{M-\mu_0}{\sqrt{\frac{s^2}{n}}}$$

magnitude

standard err



- As df grows, t(df) becomes more normal (*SS* 9.2.3).
- t(∞) = N(0, 1)
  = the Normal distribution
- Set a critical value based on the area under the t-distribution, just like we did with the normal dist. for Z
- Lookup on Table T

Critical values only.

We always do 2-tailed tests!

Notice at df = ∞, crit t = crit z

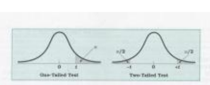


Table D.6 Percentage Points of the t Distribution (Source: The entries in this table were computed by the author.)

	Level of Significance for One-Tailed Test								
	.25	.20	.15	.10	.05	.025	.01	.005	[illegible]
	Level of Significance for Two-Tailed Test								
df	.50	.40	.30	.20	.10	.05	.02	.01	[illegible]
1	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.657	[illegible]
2	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	[illegible]
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	[illegible]
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	[illegible]
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	[illegible]
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	[illegible]
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	[illegible]
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	[illegible]
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	[illegible]
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	[illegible]
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	[illegible]
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	[illegible]
13	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	[illegible]
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	[illegible]
15	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	[illegible]
16	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	[illegible]
17	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	[illegible]
18	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	[illegible]
19	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	[illegible]
20	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	[illegible]
21	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	[illegible]
22	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	[illegible]
23	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	[illegible]
24	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	[illegible]
25	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	[illegible]
26	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	[illegible]
27	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	[illegible]
28	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	[illegible]
29	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	[illegible]
30	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	[illegible]
40	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	[illegible]
50	0.679	0.849	1.047	1.299	1.676	2.009	2.403	2.678	[illegible]
100	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	[illegible]
∞	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	[illegible]

Example - Single Sample t test

- Does Drug A raise temperatures?
- Null hypothesis: $H_0: \mu = 98.6$
- Alternative hypothesis: $H_1: \mu \neq 98.6$
- Data:

99.2, 98.7, 99.3, 97.9, 100.3, 99.4, 98.9, 99.8, 100.1

$$\bar{X} = \frac{99.2 + 98.7 + 99.3 + 97.9 + 100.3 + 99.4 + 98.9 + 99.8 + 100.1}{9} = 99.3$$

$$s = \sqrt{\frac{(99.2 - 99.3)^2 + (98.7 - 99.3)^2 + (99.3 - 99.3)^2 + \ldots + (100.1 - 99.3)}{8}} = 0.74$$

$$t_8 = \frac{\bar{X} - \mu}{s_{\bar{X}}} = \frac{\bar{X} - \mu}{s/\sqrt{n}} = \frac{99.3 - 98.6}{0.74/\sqrt{9}} = \frac{0.7}{.247} = 2.83 \qquad crit_{.05}\ t_8 = \pm 2.306$$

- Reject H_0: Patients taking Drug A have temperatures significantly above normal

Assumptions: One-sample t-test

- Independent observations (one observation doesn't influence another observation and knowing one score doesn't tell you anything about other scores)
- The scores from the populations are normally distributed

- What if the assumptions are violated?
  - Independence violated. Not good -- consider a different experiment or analysis method.
  - Normally distributed scores violated. Not a big deal, esp as sample size increases (e.g., n > 30)

Independent Samples t-test

- Is the mean in Group A different from the mean in Group B when there is no way to link up particular scores in one group with particular scores in another (the groups are independent)
- Random assignment of people to either the treatment or control conditions
- Are there gender differences in personality dimensions?
- Do CU in-state students have higher GPAs than CU out-of-state students?
- Do those who got the own-attitudes-first form of the questionnaire think about Obama differently from those who got the own-attitudes-later form of the questionnaire?
- Equal or unequal n's in the two groups
- Null hypothesis: $H_0: \mu_1 = \mu_2;\ \mu_1 - \mu_2 = 0$

Null hypothesis true: No mean difference

- Each group is actually sampled from one and the same population
- What then is the sampling distribution for each of the two group means?
  - Approximately normal
  - One true mean: μ (avg $\bar{X} = \mu$)
  - Standard error of mean $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$
    - estimated as $s_{\bar{X}} = \frac{s}{\sqrt{n}}$

Null hypothesis true: No mean difference

- We draw 2 samples from this one population, one of size n_1 and one of size n_2.
- What does the sampling distribution for the difference between the two sample means look like? (Sampling distribution of mean differences)
  - Approximately normal
  - Average $\bar{X}_1 - \bar{X}_2 = \mu - \mu = 0$
    - Since each sample mean on average equals μ
  - Standard error of mean difference

$$\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}$$

Null hypothesis true: No mean difference

- Sampling distribution of mean differences
  - Approximately normal
  - Average $\bar{X}_1 - \bar{X}_2 = 0$
    - Since each sample mean on average equals μ
  - Standard error of mean difference

$$\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}$$

    - Note parallel to standard error of mean $\sigma_{\bar{X}} = \sigma/\sqrt{n} = \sqrt{\frac{\sigma^2}{n}}$
    - How estimated?
      - We have two different sample estimates of σ

Estimating the Standard Error of the Mean Difference

$$\sigma_{\bar{X}_1-\bar{X}_2} = \sqrt{\left(\frac{\sigma^2}{n_1}+\frac{\sigma^2}{n_2}\right)}$$

$$s_{\bar{X}_1-\bar{X}_2} = \sqrt{\left(\frac{s_p^2}{n_1}+\frac{s_p^2}{n_2}\right)}$$

$$s_p^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{df_1+df_2} = \frac{SS_1+SS_2}{df_1+df_2}$$

$$If\ n_1=n_2=n, then\ s_p^2 = \frac{(n-1)s_1^2+(n-1)s_2^2}{n+n-2} = \frac{(n-1)s_1^2+(n-1)s_2^2}{2(n-1)} = \frac{s_1^2+s_2^2}{2}$$

Independent Samples t-test: Testing the null hypothesis H_0: $\mu_1 = \mu_2$

$$t_{n_1+n_2-2} = \frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{s_{\bar{X}_1-\bar{X}_2}}$$

$$= \frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{\left(\frac{s_p^2}{n_1}+\frac{s_p^2}{n_2}\right)}}$$

$$s_p^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{df_1+df_2} = \frac{SS_1+SS_2}{df_1+df_2}$$

Compare to crit $t_{n_1+n_2-2};\ \alpha = .05$

Independent t-test: Data example

	Drug	Placebo
	6	8
	3	6
	5	4
	7	7
	5	6
	4	9
	3	7
	6	5
	4	6
		5
Mean	4.78	6.30
S	1.39	1.49
SS	15.56	20.1
n	9	10

$$H_0: \mu_p = \mu_d : \mu_p - \mu_d = 0$$

$$H_1: \mu_p \neq \mu_d : \mu_p - \mu_d \neq 0$$

$$s_p^2 = \frac{(n_p-1)S_p^2+(n_d-1)S_d^2}{n_p+n_d-2} = \frac{9(1.49)^2+8(1.39)^2}{17}$$

$$= \frac{9(2.22)+8(1.90)}{17} = 2.09$$

$$= \frac{SS_1+SS_2}{df_1+df_2} = \frac{15.56+20.1}{8+9} = 2.09$$

$$t_{n_p+n_d-2} = \frac{(\bar{X}_p-\bar{X}_d)-(\mu_p-\mu_d)}{s_{\bar{X}_p-\bar{X}_d}} = \frac{(\bar{X}_p-\bar{X}_d)-(\mu_p-\mu_d)}{\sqrt{\left(\frac{s_p^2}{n_p}+\frac{s_p^2}{n_d}\right)}}$$

$$= \frac{(6.30-4.78)-(0)}{\sqrt{\left(\frac{2.09}{10}+\frac{2.09}{9}\right)}} = \frac{1.52}{.66} = 2.29$$

$$crit\ t_{17} = \pm 2.11$$

Independent t-test: Data example

	Drug	Placebo
	6	8
	3	6
	5	4
	7	7
	5	6
	4	9
	3	7
	6	5
	4	6
		5
Mean	4.78	6.30
S	1.39	1.49
SS	15.56	20.1
n	9	10

$$H_0: \mu_p = \mu_d : \mu_p - \mu_d = 0$$

$$H_1: \mu_p \neq \mu_d : \mu_p - \mu_d \neq 0$$

$$t_{n_p+n_d-2} = \frac{(\bar{X}_p-\bar{X}_d)-(\mu_p-\mu_d)}{s_{\bar{X}_p-\bar{X}_d}} = \frac{1.52}{.66} = 2.29$$

$$crit\ t_{17} = \pm 2.11$$

Reject H_0

The Placebo group has significantly higher scores than the Drug group

Assumptions: Two-sample t-test

- Independent observations (one observation doesn't influence another observation and knowing one score doesn't tell you anything about other scores)
- The two populations from which samples are drawn have equal variance
- The scores from the two populations are normally distributed

- What if the assumptions are violated?
  - Independence violated. Not good -- consider a different experiment or analysis method.
  - Equal variance violated. No big deal. Use alternative formula
  - Normally distributed scores violated. Not a big deal, esp as sample size increases (e.g., n > 30)