

Soci708 – Statistics for Sociologists

Module 7 – Inference for Distributions¹

François Nielsen

University of North Carolina
Chapel Hill

Fall 2009

¹Adapted from slides for the course Quantitative Methods in Sociology (Sociology 6Z3) taught at McMaster University by Robert Andersen (now at University of Toronto)

Outline

Introduction

Inference for the Mean of a Population

The t Distributions

The One-Sample t Confidence Interval

The One-Sample t Test

Matched Pairs t Procedures

Comparing Two Means

The Two-Sample z Statistic

The Two-Sample t Procedures

The Two-Sample t Significance Test

The Two-Sample t Confidence Interval

Topics in Comparing Distributions

Measuring Effect Size: Cohen's d

Inference for Population Spread: F Test

The Power of the Two-Sample t Test

Topics in Module 7

- ▶ TBA

The t Distributions

Review: Inference When σ is Known

- ▶ Recall the three characteristics of the sampling distribution of a sample mean (\bar{x}):
 1. It is close to normally distributed (*central limit theorem*)
 2. The mean of all the means (i.e., the mean of \bar{x}) equals the population mean μ
 3. The standard deviation of the mean is:

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

- ▶ Recall that in Module 6 we proceeded as if we knew σ

The t Distributions

Review: Inference When σ is Known

- ▶ Recall that we can standardize \bar{x} if we know σ , resulting in the one-sample z statistic:

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

Where z tells us how far the observed sample mean (\bar{x}) is from μ in the units of the standard deviation σ / \sqrt{n} of \bar{x} .

- ▶ Following from the z -statistic, we can solve for a confidence interval for μ :

$$\mu = \bar{x} \pm z^* \frac{\sigma}{\sqrt{n}}$$

Where z^* represents the critical value of z — i.e., the value of z that marks off the area under the normal curve corresponding to the specified confidence level

The t Distributions

Review: Inference When σ is Known – Find a Confidence Interval for μ

1. Calculate the point estimate for the sample mean:

$$\bar{x} = \frac{\sum x_i}{n}$$

2. Find the critical value of z^* that corresponds to the desired level of confidence
 - ▶ For example, a 95% CI requires the middle 95% of the area. Therefore, we need z^* for an area between $(1 - .95)/2 = .025$ and $1 - .05/2 = .975$. We see from Table A, that $z^* \approx 1.96$
3. Substitute the known values into the formula:

$$\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}}$$

The t Distributions

Review: Inference When σ is Known – Test a Hypothesis for μ

1. Start with an *alternative hypothesis*:

$$H_a : \mu > \mu_0$$

Remember, any direction ($\mu < \mu_0$, $\mu \neq \mu_0$, or $\mu > \mu_0$) can be specified.

2. The *null hypothesis* is then stated as the complementary event to H_a :

$$H_0 : \mu \leq \mu_0$$

3. Carry out a *test of significance* of $\bar{x} - \mu_0$:

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

High z-scores (small P-values) are evidence *against* the null hypothesis.

- ▶ We can specify an α level in advance, or simply state the *significance level* (P-value)

The t Distributions

What do we do when σ is not known?

- ▶ The problem here for real applications is that we do not know the population standard deviation σ and hence cannot determine the standard deviation of the sampling distribution
- ▶ When σ is not known, we need some way of estimating the standard deviation of \bar{x}
- ▶ We do this by using information from the sample
 - ▶ More specifically we substitute σ with s , the sample standard deviation of X
 - ▶ This results in replacing the standard deviation σ/\sqrt{n} of \bar{x} with its *standard error*:

$$SE_{\bar{x}} = \frac{s}{\sqrt{n}}$$

- ▶ The resulting test statistic is

$$t = \frac{\bar{x} - \mu}{SE_{\bar{x}}} = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

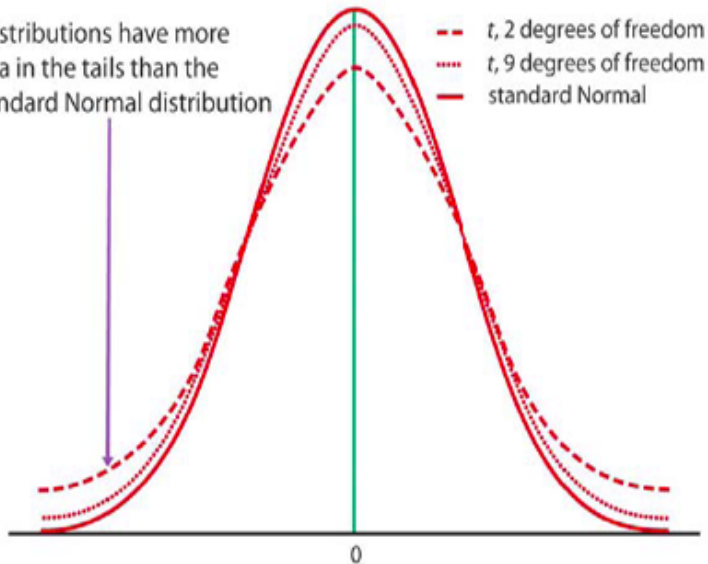
- ▶ The test statistic *is no longer normally distributed* — it follows a *Student t* distribution with $n - 1$ degrees of freedom

The t Distributions

- ▶ The t distributions are *similar in shape to the normal distribution* – they are symmetric, have a mean of zero, are single peaked and bell-shaped
- ▶ The t distributions *differ from the normal distribution in terms of spread*, however
 - ▶ Because we substitute s for the fixed parameter σ , we add more variation into the statistic (in other words, less certainty). This results in the standard deviation of the t distribution being larger than the standard deviation of the normal distribution
 - ▶ As the sample size n (and degrees of freedom, $df = n - 1$) increases the t distribution approaches the $N(0, 1)$
 - ▶ In other words, with large sample sizes, t and z give nearly identical results
 - ▶ This can be seen in IPS6e Table D (see bottom row)

The t -Distributions

t distributions have more area in the tails than the standard Normal distribution



The t Distributions

“Student” Strikes Again! Discovery of the t Distributions



- ▶ William Gosset published the t distribution under the pseudonym *Student* while employed by the Guinness brewery in Dublin



The t Distributions

More on “Student”: Origin of the t -test in industrial quality control²

*The t statistic was introduced in 1908 by William Sealy Gosset, a statistician working for the Guinness brewery in Dublin, Ireland (“Student” was his pen name). Gosset had been hired due to Claude Guinness’s innovative policy of recruiting the best graduates from Oxford and Cambridge to apply biochemistry and statistics to Guinness’ industrial processes. Gosset devised the t -test as a way to cheaply monitor the quality of beer. He published the test in *Biometrika* in 1908, but was forced to use a pen name by his employer, who regarded the fact that they were using statistics as a trade secret. In fact, Gosset’s identity was known to fellow statisticians.*

²From Wikipedia article *Student’s t -test* accessed 30 Oct 2008

The One-Sample t Confidence Interval

- ▶ Procedures for confidence intervals and significance tests for a population mean follow exactly the same structure as when using the normal distribution
- ▶ All we do now, then, is use s in place of σ and the t -statistic instead of the z -statistic
- ▶ So, if we want a *confidence interval for a population mean* we use the following formula:

$$\bar{x} \pm t^* SE_{\bar{x}} = \bar{x} \pm t^* \frac{s}{\sqrt{n}}$$

where $SE_{\bar{x}} = s/\sqrt{n}$

The One-Sample t Confidence Interval

- ▶ The quantity $SE_{\bar{x}} = s/\sqrt{n}$ is called the *standard error* of the mean; it is an estimate (from sample data) of $\sigma_{\bar{x}} = \sigma/\sqrt{n}$, the standard deviation of the sampling distribution of \bar{x}
- ▶ The quantity t^*s/\sqrt{n} in the formula for the confidence interval is called the *margin of error*, denoted m
- ▶ Here t^* is the upper $(1 - C)/2$ critical value for the $t(n - 1)$ distribution; recall the C is the level of confidence we choose
- ▶ The value of t^* corresponding to a given C and given $df = n - 1$ can be found in Table D or from software

Table D

When σ is unknown, we use a t distribution with " $n-1$ " degrees of freedom (df).

Table D shows the z -values and t -values corresponding to landmark P -values/ confidence levels.

When σ is known, we use the normal distribution and the standardized z -value.

df	Upper tail probability p											
	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	127.3	318.3	636.6
2	0.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	0.765	0.978	1.250	1.636	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.908	7.171	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	2.757	3.365	4.037	4.773	5.893	6.869
6	0.718	0.908	1.134	1.440	1.895	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.885	2.365	2.517	2.998	3.499	4.029	4.783	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.449	2.896	3.353	3.833	4.501	5.041
9	0.705	0.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.146	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.023	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.426	3.910	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.283	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.762	2.149	2.272	2.638	2.977	3.328	3.787	4.140
15	0.691	0.866	1.074	1.341	1.756	2.140	2.263	2.628	2.967	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	1.751	2.133	2.256	2.621	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.747	2.127	2.250	2.615	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.330	1.744	2.122	2.245	2.610	2.878	3.197	3.611	3.922
19	0.688	0.861	1.066	1.327	1.741	2.118	2.241	2.605	2.861	3.174	3.579	3.883
20	0.687	0.860	1.064	1.324	1.738	2.114	2.237	2.601	2.845	3.151	3.551	3.850
21	0.686	0.859	1.063	1.321	1.735	2.110	2.233	2.597	2.831	3.133	3.527	3.819
22	0.686	0.858	1.061	1.321	1.732	2.107	2.230	2.593	2.819	3.119	3.503	3.792
23	0.685	0.858	1.060	1.319	1.729	2.104	2.227	2.590	2.807	3.104	3.485	3.768
24	0.685	0.857	1.059	1.318	1.727	2.101	2.224	2.587	2.797	3.091	3.467	3.745
25	0.684	0.856	1.058	1.316	1.725	2.100	2.221	2.585	2.787	3.078	3.450	3.725
26	0.684	0.856	1.058	1.315	1.724	2.099	2.219	2.583	2.779	3.067	3.435	3.707
27	0.684	0.855	1.057	1.314	1.723	2.098	2.218	2.582	2.772	3.057	3.421	3.690
28	0.683	0.855	1.056	1.313	1.721	2.098	2.217	2.581	2.765	3.047	3.408	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.150	2.463	2.756	3.038	3.396	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.147	2.457	2.750	3.030	3.385	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.123	2.423	2.704	2.971	3.307	3.551
50	0.679	0.849	1.047	1.299	1.676	2.009	2.109	2.403	2.678	2.937	3.261	3.496
60	0.679	0.848	1.045	1.296	1.671	2.000	2.099	2.390	2.660	2.915	3.232	3.460
80	0.678	0.846	1.043	1.292	1.664	1.990	2.088	2.374	2.639	2.887	3.195	3.416
100	0.677	0.845	1.042	1.290	1.660	1.984	2.081	2.364	2.626	2.871	3.171	3.390
1000	0.675	0.842	1.037	1.281	1.646	1.962	2.056	2.330	2.581	2.815	3.096	3.300
z^*	0.674	0.841	1.036	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.091	3.291
	50%	60%	70%	80%	90%	95%	96%	98%	99%	99.5%	99.8%	99.9%
	Confidence level C											

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

Table A vs. Table D

Table A gives the area to the LEFT of hundreds of z-values.

It should only be used for Normal distributions.

TABLE A Standard normal probabilities

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
-3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
-3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0006	.0005	.0005
-3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
-3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
-2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
-2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
-2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
-2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
-2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
-2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
-2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
-2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110

(...)

Table D t distribution critical values

df	Upper tail probability p											
	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	127.3	318.3	636.6
2	0.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	0.766	0.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
(...)												
100	0.677	0.845	1.042	1.290	1.660	1.984	2.081	2.364	2.626	2.871	3.174	3.390
1000	0.675	0.842	1.037	1.282	1.646	1.962	2.056	2.330	2.581	2.813	3.068	3.300
z*	0.674	0.841	1.036	1.282	1.645	1.960	2.051	2.326	2.576	2.807	3.061	3.291
Confidence level C												
	50%	60%	70%	80%	90%	95%	96%	98%	99%	99.5%	99.8%	99.9%

Table D gives the area to the RIGHT of a dozen t or z-values.

It can be used for t distributions of a given df and for the Normal distribution.

Table D also gives the middle area under a t or normal distribution comprised between the negative and positive value of t or z.

The One-Sample t Confidence Interval

Assumptions in Using the t -Distributions

- ▶ The data are from a *simple random sample*
 - ▶ Typically we relax this assumption to include samples that are approximate to a simple random sample
- ▶ The *population has a normal distribution* with mean μ and standard deviation σ
 - ▶ We relax this assumption and assume only that the distribution is symmetric, single peaked and does not have serious outliers
 - ▶ Like the mean, t -procedures are strongly influenced by outliers
 - ▶ The larger the sample, the more relaxed we get:
 - ▶ *Small samples of $n < 15$* : Sample data should be close to normally distributed and have no serious outliers
 - ▶ *Samples of $15 \leq n < 40$* : Use if no outliers nor strong skewness (normality not necessary)
 - ▶ *Large samples of $n \geq 40$* : Can be used even if sample distribution is skewed

The One-Sample t Confidence Interval



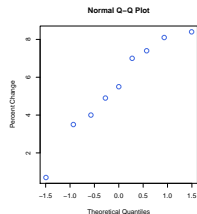
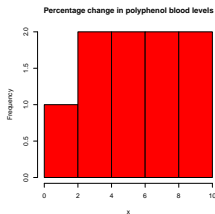
The One-Sample t Confidence Interval

Example: Red Wine, in Moderation?

- ▶ Drinking red wine in moderation may protect against heart attacks. The polyphenols it contains act on blood cholesterol, likely helping to prevent heart attacks.
- ▶ To see if moderate red wine consumption increases the average blood level of polyphenols, a group of nine randomly selected healthy men were assigned to drink half a bottle of red wine daily for two weeks. Their blood polyphenol levels were assessed before and after the study, and the percent change is presented here:

0.7 3.5 4 4.9 5.5 7 7.4 8.1 8.4

0 | 7
2 | 5
4 | 095
6 | 04
8 | 14



There is a low value, but overall the data can be considered reasonably normal.

The One-Sample t Confidence Interval

Example: Red Wine, in Moderation?

- ▶ What is the 95% confidence interval for the average percent change?
- ▶ Sample average $\bar{x} = 5.5$; $s = 2.517$; $df = n - 1 = 8$
- ▶ The sampling distribution is a t distribution with $n - 1$ degrees of freedom.
 - ▶ For $df = 8$ and $C = 95\%$, $t^* = 2.306$.
- ▶ The margin of error m is

$$m = t^* s / \sqrt{n} = 2.306 * 2.517 / \sqrt{9} \approx 1.93.$$

- ▶ With 95% confidence, the population average percent increase in polyphenol blood levels of healthy men drinking half a bottle of red wine daily is between 3.6% and 7.6%.
 - ▶ *Important:* The confidence interval shows how large the increase is, but not if it can have an impact on men's health.

The One-Sample t Confidence Interval

R and in Stata Calculations for Red Wine Example

```
> # in R
> x <-c(0.7, 3.5, 4, 4.9, 5.5,
        7, 7.4, 8.1, 8.4)
> mean(x)
[1] 5.5
> sd(x)
[1] 2.516943
> # t(8)* for 95% confidence
> qt(0.975, 8)
[1] 2.306004
> # Margin of error
> 2.306004*(2.516943/sqrt(9))
[1] 1.934694
> # CI lower and upper bounds
> 5.5 - 1.934694
[1] 3.565306
> 5.5 + 1.934694
[1] 7.434694

. * in Stata
. * change for the red wine problem
. display invttail(999, 0.025)
1.9623415
. * Margin of error
. display invttail(999, 0.025)
*(4.5/sqrt(1000))
.27924609
. * CI lower and upper bounds
. display 12.5 - 0.27924609
12.220754
. display 12.5 + 0.27924609
12.779246
```

The One-Sample t Confidence Interval

Normal Distribution versus t -Distributions

- ▶ For the red wine example, the 95% CIs using the t -distribution and the normal distribution would end up substantially different
 - ▶ The critical value of t for $n = 9$ (8 df) is 2.306 — quite larger than the critical value for z of 1.960 for a 95% CI using the normal distribution
- ▶ Let's compare the results for a 95% CI for the t distribution and normal distribution for this same example ($n = 9$, $df = 8$):

$$\begin{aligned}\bar{x} \pm t^* \frac{s}{\sqrt{n}} &= 5.5 \pm 2.306 \frac{2.517}{\sqrt{9}} & \bar{x} \pm z^* \frac{s}{\sqrt{n}} &= 5.5 \pm 1.960 \frac{2.517}{\sqrt{9}} \\ &= 5.5 \pm 1.935 & &= 5.5 \pm 1.644 \\ &= 3.6 \text{ to } 7.6 & &= 3.9 \text{ to } 7.1\end{aligned}$$

- ▶ We see here that the CI is larger for the t distribution
 - ▶ The difference between t^* and z^* gets bigger the smaller the sample size

The One-Sample t Test

- ▶ The one-sample t statistic is as follows:

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

- ▶ Notice that the only thing different here from the z-test is that we use s in place of σ
- ▶ Since each sample will have a different standard deviation s , we incorporate this added uncertainty into our estimation of the test statistic.
 - ▶ The t statistic has a different degrees of freedom for each sample size, resulting in greater confidence as the sample size increases. The degrees of freedom for t are $n - 1$.
 - ▶ There is a separate distribution for each n
- ▶ We write the t distribution with k degrees of freedom as $t(k)$. See Table D in IPS6e for critical values of t

The One-Sample t Test

Degrees of Freedom (1)

- ▶ The degrees of freedom is usually denoted by df
- ▶ This is the number of values (observations) that are allowed to vary when computing a statistic
 - ▶ In other words, it is the number of independent pieces of information that are used in calculating the statistic
 - ▶ In practical terms, it is almost synonymous with sample size
- ▶ The df of a t distribution is inherited from that of the sample variance s^2 — the square root s of which is in the denominator — which has $n - 1$ df :

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1}$$

The One-Sample t Test

Degrees of Freedom (2)

- ▶ Recall that to calculate the variance s^2 , we start by calculating the mean ($\sum x_i/n$). Next we find the sum of the squared deviations from that mean:

$$\sum (x_i - \bar{x})^2$$

- ▶ There are n squared deviations but *only* $(n - 1)$ of them are free to take on any value whatsoever
 - ▶ The last squared deviation from the mean includes the *one value of X* that ensures that $\sum x_i/n$ equals the mean of the sample.
- ▶ For these reasons, the variance s^2 — and the corresponding t distribution — is said to have only $(n - 1)$ degrees of freedom

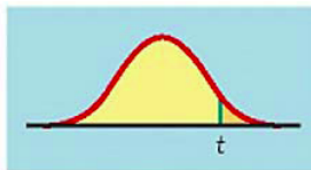
The One-Sample t Test

- ▶ For a t test, we specify the alternative hypothesis and the null hypothesis in the same manner as for a significance tests based on the z -statistic
- ▶ As before, we can specify one of three different alternative hypotheses:
 - ▶ $H_a : \mu > \mu_0$ (we look at the area to the right of t)
 - ▶ $H_a : \mu < \mu_0$ (we look at the area to the left of t)
 - ▶ $H_a : \mu \neq \mu_0$ (we look at the area more extreme than $|t|$)
- ▶ The null hypothesis H_0 is then the *complement* of H_a and always includes equality $\mu = \mu_0$
- ▶ The one-sample t -statistic with $n - 1$ df is:

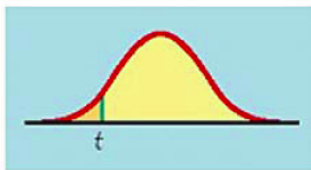
$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{x}}$$

The One-Sample t Test

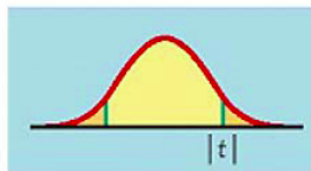
$$H_a: \mu > \mu_0$$



$$H_a: \mu < \mu_0$$



$$H_a: \mu \neq \mu_0$$



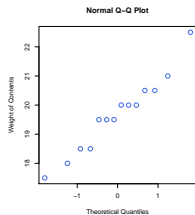
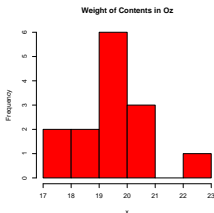
The One-Sample t Test

Example: Underweight Tomatoes? (Koopmans 1987, p. 307)

- ▶ A consumer group suspects that the average weight of canned tomatoes produced by a cannery is less than the advertised weight of 20 oz. The group purchases 14 cans from various stores and weighted the contents. The weights were as follows (to the nearest half oz).

20.5	18.5	20.0	19.5	19.5	21.0	17.5
22.5	20.0	19.5	18.5	20.0	18.0	20.5

16		5
18		055555
20		000550
22		5



The data appear reasonably normal.

The One-Sample t Test

Example: Underweight Tomatoes? (Koopmans 1987, p. 307)

- ▶ We want to test the hypothesis that the mean weight μ is less than $\mu_0 = 20$ oz at the $\alpha = .05$ significance level
- ▶ We know or calculate the following information:

$n = 14$ (sample size)

$\bar{x} = 19.679$ (mean weight)

$s = 1.295$ (standard deviation)

1. We start by stating the alternative hypothesis, which is that the weight of the contents is *less than* 20:

$$H_a : \mu < 20$$

Thus it is a *one-sided* hypothesis.

The One-Sample t Test

Example: Underweight Tomatoes?

2. We state the (complementary) null hypothesis which in this case is that mean weight is 20 oz (or more):

$$H_0 : \mu \geq 20$$

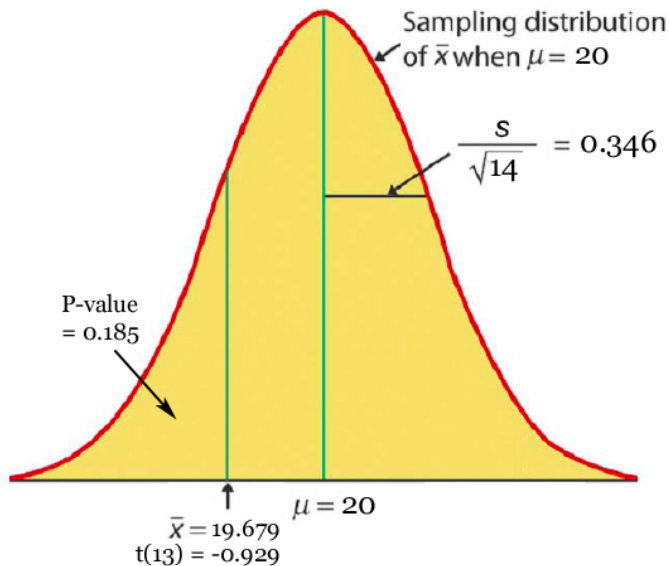
3. We now carry out a t -test to see what proportion of samples would give an outcome as extreme as ours (19.679) if the null hypothesis ($\mu = 20$) is correct. The t -statistic is calculated as:

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{19.679 - 20}{1.295/\sqrt{14}} = -0.929$$

4. In Table D we see that for $df = 13$ the P-value for 0.929 is between .15 and .20. Software gives .185. Thus we cannot reject at $\alpha = .05$ the null hypothesis that mean contents weight is 20 oz or more. In other words mean weight is *not significantly less* than 20 oz.

The One-Sample t Test

Example: Underweight Tomatoes?



The One-Sample t Test

One-sample t-tests in R: Underweight Tomatoes?

```
> x <- c(20.5, 18.5, 20.0, 19.5, 19.5, 21.0, 17.5,
         22.5, 20.0, 19.5, 18.5, 20.0, 18.0, 20.5)
> mean(x)
[1] 19.67857
> sd(x)
[1] 1.295067
> # t-statistic
> (19.67857-20)/(1.295067/sqrt(14))
[1] -0.9286631
> # P-value
> pt(-0.9286631, 13)
[1] 0.1849932
> # Or using t.test procedure
> t.test(x, alternative="less", mu=20, conf.level=.95)
      One Sample t-test
```

```
data: x
t = -0.9287, df = 13, p-value = 0.185
alternative hypothesis: true mean is less than 20
95 percent confidence interval:
 -Inf 20.29153
```

```
> # other tests: alternative="greater"; alternative="two.sided"
```


Matched Pairs t Procedures

A Special Kind of One-sample t -test

- ▶ So far we have only examined one-sample t -tests for a single population mean
- ▶ A matched pairs test is a special kind of one-sample test that *tests for difference in a particular variable measured twice for a single observation*
 - ▶ For example, a test for difference in salary between husband and wives is a matched pairs design — the husbands and wives are not independent observations. In this case “married couples” are the observations
 - ▶ Another example of a matched pairs design is a test for a change over time for individuals – i.e., information is gathered from people at two points in time, and we test for difference over time
- ▶ The matched pairs test is simply a single sample test of the mean of the *difference* of the two values for the pair

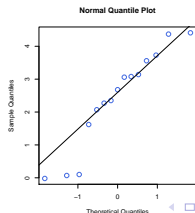
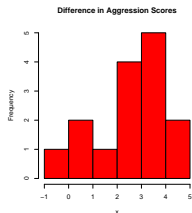
Matched Pairs t Procedures

Example: Aggressive Behavior of Patients During Full Moon (IPS6e)

- ▶ We have data on number of aggressive acts by dementia patients during a period when the moon is not full and the same patients during a full moon period. The data looks like the following:

patient	aggmoon	aggother	$X=\text{aggdiff}$
1	3.33	0.27	3.06
2	3.67	0.59	3.08
3	2.67	0.32	2.35
...
15	2.00	0.38	1.62

```
-0 | 0
0 | 11
1 | 6
2 | 1347
3 | 11167
4 | 44
```



The data appear reasonably normal although we note 3 patients with low values of x .

Matched Pairs t Procedures

Example: Aggressive Behavior of Patients During Full Moon

- ▶ The alternative hypothesis is that the level of aggression of dementia patients during full moon is *different* than at other periods. Thus the hypotheses are set up as

$$H_a : \mu \neq 0$$

$$H_0 : \mu = 0$$

- ▶ We have or we calculate the following information:

$$n = 15$$

$$\bar{x} = 2.433 \quad (\text{mean difference in aggression})$$

$$s = 1.460 \quad (s \text{ for difference in aggression})$$

$$\mu_0 = 0$$

- ▶ The one sample $t(n-1)$ test statistic for the mean difference is:

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

Matched Pairs t Procedures

Example: Aggressive Behavior of Patients During Full Moon

- ▶ Substituting in the known values we get:

$$t = \frac{\bar{x} - 0}{s/\sqrt{n}} = \frac{2.433}{1.460/\sqrt{15}} = 6.452$$

- ▶ We can see from Table D that the upper tail p-value for $t(14) = 6.452$ is much smaller than .0005
- ▶ We need to double this one-tailed p-value (as this is a two-tailed test), calculating later that the two-sided p-value is $1.518152e-05$ or .00001518. Thus we are quite confident that we can reject the hypothesis of equality
- ▶ We conclude, then, that there is a significant difference in aggression between the full moon and other periods

Matched Pairs t Procedures

R Example: Aggressive Behavior of Patients During Full Moon (IPS6e)

```
> Fullmoon <- read.table("D:/soci708/
  data/data_from_Moore_2009/PC-Text/
  ch07/ta07_002.txt", header = TRUE)
> attach(Fullmoon)
> Fullmoon
  patient aggmoon aggother aggdiff
1         1     3.33     0.27     3.06
2         2     3.67     0.59     3.08
3         3     2.67     0.32     2.35
4         4     3.33     0.19     3.14
5         5     3.33     1.26     2.07
6         6     3.67     0.11     3.56
7         7     4.67     0.30     4.37
8         8     2.67     0.40     2.27
9         9     6.00     1.59     4.41
10        10     4.33     0.60     3.73
11        11     3.33     0.65     2.68
12        12     0.67     0.69    -0.02
13        13     1.33     1.26     0.07
14        14     0.33     0.23     0.10
15        15     2.00     0.38     1.62

> xbar <- mean(aggdiff)
> xbar
[1] 2.432667
> s <- sd(aggdiff)
> s
[1] 1.460320
> n <- 15
> t <- (xbar-0)/(s/sqrt(n))
> t
[1] 6.451789
> pval <- 2*(1-pt(t,14))
> pval
[1] 1.518152e-05
> # Or use t.test function
> t.test(aggdiff, alternative=
  "two.sided", mu=0)
...
t = 6.4518, df = 14,
  p-value = 1.518e-05
```

Matched Pairs t Procedures

R Example: Aggressive Behavior of Patients During Full Moon (IPS6e)

```
> # In R
> # Dataset Fullmoon has been read in
> # as in previous slide...
> # One can do 1-sample t-test
> # on aggdiff...
> t.test(aggdiff, alternative="two.sided",
          mu=0)
```

One Sample t-test

```
data: aggdiff
t = 6.4518, df = 14, p-value = 1.518e-05
alternative hypothesis: true mean is not
equal to 0
95 percent confidence interval:
 1.623968 3.241365
sample estimates:
mean of x
 2.432667
```

```
> # Or specify paired=TRUE for t-test with
          two original variables
> t.test(aggmoon, aggother, paired=TRUE,
          alternative="two.sided", mu=0)
```

Paired t-test

```
data: aggmoon and aggother
t = 6.4518, df = 14, p-value = 1.518e-05
alternative hypothesis: true difference
in means is not equal to 0
95 percent confidence interval:
 1.623968 3.241365
sample estimates:
mean of the differences
 2.432667
```

Matched Pairs t Procedures

Stata Example: Aggressive Behavior of Patients During Full Moon (IPS6e)

```
* open ta07_002.xls in Excel, copy data
* open Data Editor in Stata, paste data
* click Preserve then CLOSE Data Editor
* to save as *.dta file use File/Save
. list ... (output not shown)
. stem aggdiff ...

-0** | 02
0** | 07,10
0** |
1** |
1** | 62
2** | 07,27,35
2** | 68
3** | 06,08,14
3** | 56,73
4** | 37,41

. * first do it by hand for practice
. su aggdiff
```

```
. display 2.432667/(1.46032/sqrt(15))
6.4517906

. display 2*ttail(14, 6.4517906)
.00001518

. * now the easy way!
. ttest aggdiff=0
... (output not shown)
```

Variable	Obs	Mean	Std. Dev.	Min	Max
-----+-----					
aggdiff	15	2.432667	1.46032	-.02	4.41

The Two-Sample z Statistic

- ▶ To be added: Discussion of two-sample z -statistic corresponding to pp. 447–451 in IPS6e, culminating in the formula (p. 450):

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

The Two-Sample t Procedures

- ▶ Add general discussion of 2-sample t -statistic here corresponding to IPS6e pp. 450–451, culminating in the formula

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

- ▶ Need for approximation for the degrees of freedom of test statistic

The Two-Sample t Significance Test

- ▶ Often we want to *compare the means from two different populations or subgroups*
 - ▶ For example, testing whether men have higher incomes than women. In this case, women and men constitute separate populations
- ▶ Such comparisons require *two sample t -test procedures*
- ▶ A *confidence interval* is found as follows:

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

- ▶ Here the subscripts are used to denote which sample the numbers come from. So, we have two samples: one labeled “1” and the other labeled “2”
- ▶ The degrees of freedom for t^* is the smaller of $n_1 - 1$ and $n_2 - 1$

The Two-Sample t Significance Test

- ▶ A two-sample t -statistic for a significance test for the hypothesis $H_0 : \mu_1 = \mu_2$ is found as follows:

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

- ▶ Once again the subscripts denote which sample the numbers come from: one sample is labeled “1”; the other labeled “2”
- ▶ As was the case for the critical value of t for the confidence interval, the degrees of freedom for t is the smaller of $n_1 - 1$ and $n_2 - 1$ ³

³Software uses an approximation formula for df ; estimated df is between n_1 and n_2 and not necessarily an integer

The Two-Sample t Significance Test

Assumptions for Comparing Two Means

1. Both samples should be *SRS of two distinct populations or subgroups*
 - ▶ The samples must be independent of each other
 - ▶ In practice, subsamples of a single SRS satisfy this condition
2. Both populations *must be normally distributed*.
 - ▶ We cannot know whether the populations are normally distributed, but just as with the one sample tests and confidence intervals for a mean, we can (and should) *look at the sample distribution* for signs of nonnormality
 - ▶ If the population distributions differ, larger sample sizes are needed
3. Finally, the *two sample tests are most robust when the two samples are the same size*

The Two-Sample t Significance Test

Example: Proportion of women in State cabinets

- ▶ Suppose we want to test whether the proportion of women in State cabinets is higher in states with Democratic governors (population 1) than in states with Republican governors (population 2)
 - ▶ The alternative hypothesis is: $H_a : \mu_1 > \mu_2$
 - ▶ The null hypothesis is: $H_0 : \mu_1 \leq \mu_2$
- ▶ We chose the $\alpha = .05$ level
- ▶ We have the following data from the two SRS (sets of states with cabinets):

```
> Woprops[1:5,]
  state. demo womno cabno woprop
1     NJ    0      6    20  0.300
2     IL    0      7    27  0.259
3     IA    0      5    20  0.250
4     ME    0      5    20  0.250
5     WI    0      5    20  0.250
```

...

The Two-Sample t Significance Test

Example: Proportion of women in State cabinets

- ▶ The following statistics were calculated from the samples:

Democrat

$$\bar{x}_1 = 0.239864$$

$$s_1 = 0.135046$$

$$n_1 = 22$$

Republican

$$\bar{x}_2 = 0.171235$$

$$s_2 = 0.08224$$

$$n_2 = 17$$

- ▶ Substituting these values into the equation gives the t -statistic with $(n_2 - 1 = 17 - 1 = 16)$ degrees of freedom:

$$\begin{aligned} t &= \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{.239864 - .171235}{\sqrt{\frac{0.135046^2}{22} + \frac{0.08224^2}{17}}} \\ &= \frac{0.068629}{0.03502602} \\ &= 1.959372 \end{aligned}$$

The Two-Sample t Significance Test

Example: Proportion of women in State cabinets

- ▶ The t -statistic 1.959 with 16 df:
 - ▶ Is more than 1.746, one-sided critical value of $t(16)$ for $\alpha = .05$
 - ▶ Has one-sided P-value 0.0339, which is less than .05

Thus *we can reject the null hypothesis* that the proportion of women is *the same* in states with Democratic and Republican governors, against the alternative that it is *higher* in states with Democratic governors

- ▶ On the other hand if we had carried out a two-sided test, the corresponding p-value would have been $2 \times 0.0339 = 0.0677$.
 - ▶ We would conclude that *we cannot reject the null hypothesis* that the proportion of women is *the same* in Democratic and Republican-governed states against the alternative that they are *different*
- ▶ In the two-sided context we can also construct a confidence interval for the difference in the proportion of women in the cabinets Democratic and Republican governors

The Two-Sample t Significance Test

R Example: Proportion of women in State cabinets – Two-sided test

```
> library foreign # provides the read.dta procedure
> Woprops<-read.dta("D:/soci708/data/woprops.dta")
> attach(Woprops)
> t.test(woprop~demo, alternative="two.sided", conf.level=.95,
         var.equal=FALSE)
```

Welch Two Sample t-test

```
data: woprop by demo
t = -1.9594, df = 35.317, p-value = 0.058
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.139712197  0.002455513
sample estimates:
mean in group 0 mean in group 1
 0.1712353      0.2398636

> # States with democratic governor do not have a significantly
  different proportion of women cabinet members (2-sided)
```

The Two-Sample t Significance Test

R Example: Proportion of women in State cabinets – One-sided test

```
> library foreign # provides the read.dta procedure
> Woprops<-read.dta("D:/soci708/data/woprops.dta")
> attach(Woprops)
> t.test(woprop~demo, alternative="less", conf.level=.95,
          var.equal=FALSE)
```

Welch Two Sample t-test

```
data: woprop by demo
t = -1.9594, df = 35.317, p-value = 0.029
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
 -Inf -0.009463723
sample estimates:
mean in group 0 mean in group 1
 0.1712353      0.2398636

> # Now states with democratic governor have a significantly
  greater proportion of women cabinet members (1-sided)
```

The Two-Sample t Significance Test

Stata Example: Proportion of women in State cabinets – First, Look at Data

```
. stem woprop if demo
```

```
0** | 00
0** | 67
1** | 00,30
1** | 50,67,82,88
2** | 00,00,22,31,40
2** | 61,67,73,86
3** | 00,33
3** |
4** | 00,44
4** |
5** |
5** |
6** | 36
```

```
. by demo: su woprop
```

```
-> demo = 0
```

Variable	Obs	Mean	Std. Dev.	Min	Max
woprop	17	.1712353	.0822401	.048	.3

```
-> demo = 1
```

Variable	Obs	Mean	Std. Dev.	Min	Max
woprop	22	.2398636	.1350461	0	.636

```
. stem woprop if !demo
```

```
0** | 48
0** | 67,71,77
0** |
1** | 05,05,11
1** |
1** |
1** | 67
1** | 85
2** |
2** | 22,22,22
2** | 50,50,50,59
2** |
2** |
3** | 00
```

The Two-Sample t Significance Test

Stata Example: Proportion of women in State cabinets – Equal variances (default)

```
. ttest wopro, by(demo)
```

```
Two-sample t test with equal variances
```

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
0	17	.1712353	.0199462	.0822401	.1289513	.2135193
1	22	.2398636	.0287919	.1350461	.1799875	.2997397
combined	39	.2099487	.0190242	.1188063	.1714362	.2484613
diff		-.0686283	.0372071		-.144017	.0067604

```
diff = mean(0) - mean(1)                                t = -1.8445  
Ho: diff = 0                                           degrees of freedom = 37
```

```
Ha: diff < 0                                           Ha: diff != 0                                           Ha: diff > 0  
Pr(T < t) = 0.0366                                     Pr(|T| > |t|) = 0.0731                                   Pr(T > t) = 0.9634
```

```
. * contrary to R examples, Stata assumes equal variances by default  
. * df are calculated as (17 - 1) + (22 - 1) = 37
```

The Two-Sample t Significance Test

The Two-Sample t Significance Test

```
. ttest wopro, by(demo) unequal
```

```
Two-sample t test with unequal variances
```

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
0	17	.1712353	.0199462	.0822401	.1289513	.2135193
1	22	.2398636	.0287919	.1350461	.1799875	.2997397
combined	39	.2099487	.0190242	.1188063	.1714362	.2484613
diff		-.0686283	.0350261		-.1397122	.0024555

```
diff = mean(0) - mean(1)                                t = -1.9594  
Ho: diff = 0                                           Satterthwaite's degrees of freedom = 35.3172
```

```
Ha: diff < 0                                           Ha: diff != 0                                           Ha: diff > 0  
Pr(T < t) = 0.0290                                     Pr(|T| > |t|) = 0.0580                                   Pr(T > t) = 0.9710
```

- . * same as in R examples, assume unequal variances (option unequal)
- . * note p-values smaller because assumption of unequal variances
- . * seems more realistic in this situation (compare the sd's)

The Two-Sample t Confidence Interval

- ▶ Add discussion of 2-sample CI here culminating in formula (p. 454)

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

Optional Topics

- ▶ These topics will be added later

Measuring Effect Size: Cohen's d

- ▶ Once a *statistically* significant difference is found (between the sample mean and a hypothetical mean or between two sample means, one wants to evaluate how *substantively* “big” the difference is
- ▶ This can be done using Cohen's d : divide the estimated difference by the standard deviation of the population ...

Inference for Population Spread: F Test

- ▶ Discuss this topic here (pp. 473–477) ...

The Power of the Two-Sample t Test

- ▶ Discuss this topic here (pp. 477–479)
- ▶ Note how the noncentrality parameter δ is, in fact, related to Cohen's d . . .