Lecture 2
Comparing two populations – ttests and rank tests

- The *t* distribution
- **Matched pairs *t*** procedures
- Sign test and signed Wilcoxon test
- Two sample t-test and Wilcoxon-Mann Whitney test

The *t* distribution:

The goal is to estimate or test for unknown ***μ*** in situation when ***σ*** is also unknown (but not searched for).

Solution: estimate ***σ*** by ***s*** and use intelligently in formulas.

Challenge: the distribution of the test statistics will change.

Sampling – Normal Population, Unknown Standard Deviation

- Suppose an SRS $X_1, \ldots, X_n$ is selected from a normally distributed population with mean μ and standard deviation σ.
- Assume that μ and σ are both unknown.
- We know that $\bar{X} \sim N\left(\mu, \sigma/\sqrt{n}\right)$
- When σ is unknown, we estimate its value with the sample standard deviation *s*.

Sampling – Normal Population, Unknown Standard Deviation

- The standard deviation of $\bar{X}$ can be estimated by

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

- This quantity is called the standard error of the sample mean.
- The test statistic (appropriately standardized sample mean) will no longer be normally distributed when we use the standard error.
- The test statistic will have a new distribution, called the *t* (or Student's *t*) distribution.

The t-test Statistics and Distribution

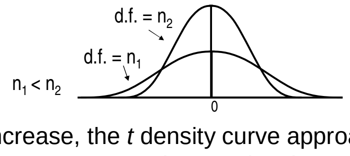
- Suppose that an SRS of size *n* is drawn from an $N(\mu, \sigma)$ population. Then the one-sample *t*-statistic

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

  has the *t* distribution with $n - 1$ degrees of freedom.
- There is a different *t* distribution for each sample size.
- The degrees of freedom for the *t*-statistic "come" from the sample standard deviation *s*.
- The density curve of a *t* distribution with *k* degrees of freedom is symmetric about 0 and bell-shaped.

The *t* Distribution

- The higher the degrees of freedom (df) are, the narrower the spread of the *t* distribution.



- As df increase, the *t* density curve approach the $N(0, 1)$ curve more and more closely.
- Generally it is more spread than normal, especially if df small.

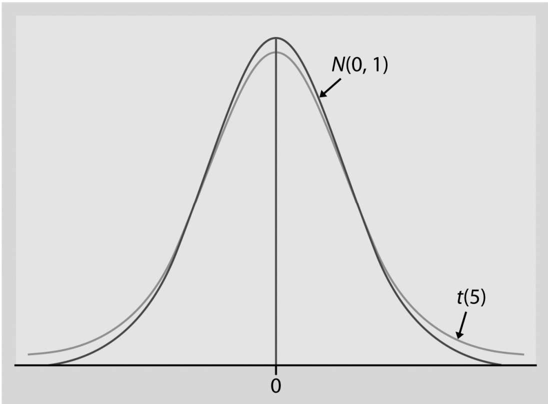


TABLE D *t* distribution critical values

	Upper tail probability *p*											
df	.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	1.638	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
8	0.706	0.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	0.703	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	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878	3.197	3.611	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861	3.174	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845	3.153	3.552	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831	3.135	3.527	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819	3.119	3.505	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807	3.104	3.485	3.768
24	0.685	0.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797	3.091	3.467	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787	3.078	3.450	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.162	2.479	2.779	3.067	3.435	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.158	2.473	2.771	3.057	3.421	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.154	2.467	2.763	3.047	3.408	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.150	2.462	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	3.174	3.390
1000	0.675	0.842	1.037	1.282	1.646	1.962	2.056	2.330	2.581	2.813	3.098	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*											

One-sample *t* Confidence Interval

- Suppose a SRS of size *n* is drawn from population having unknown mean μ. A level C confidence interval for μ is

$$\bar{x} \pm t^* \frac{s}{\sqrt{n}}, \quad \text{or} \quad \left[\bar{x} - t^* \frac{s}{\sqrt{n}},\ \bar{x} + t^* \frac{s}{\sqrt{n}}\right]$$

  Here *t** is the value for the *t* density curve with df=n-1. The area between –*t** and *t** is C.
- The interval is exact for normal population and approximately correct for large *n* in other cases.

Example

- From running production of corn soy blend we take a sample to measure content of vit. C. Results are:

  26 31 23 22 11 22 14 31.
- Find 95% confidence interval for content of vitamin C in this production.
- Give the margin of error.

Solution:

One-Sample *t* Test

- Suppose that an SRS of size *n* is drawn from a population having unknown mean μ.
- To test the hypothesis H_0: $\mu = \mu_0$ based on an SRS of size *n*, compute the one-sample t statistic

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

- Note the standard error in the denominator.

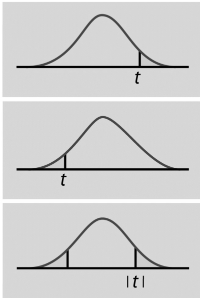
One-Sample *t* Test

- In terms of a random variable T with $t(n-1)$ distribution, the P-value for a test of H_0: $\mu = \mu_0$ against…

  H_a: $\mu > \mu_0$ is $P(T \geq t)$

  H_a: $\mu < \mu_0$ is $P(T \leq t)$

  H_a: $\mu \neq \mu_0$ is $2P(T \geq |t|)$



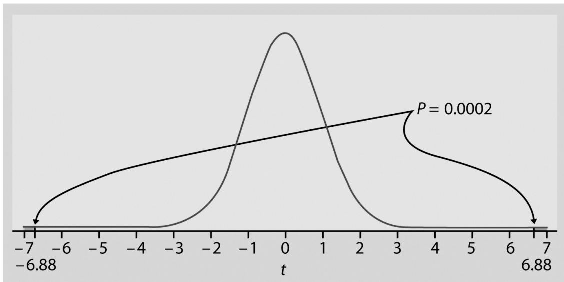
P-values

These P-values are exact if the population distribution is normal and are approximately correct for large *n* in other cases.

Example (vit. C continued):

Test whether vit. C conforms to specifications.

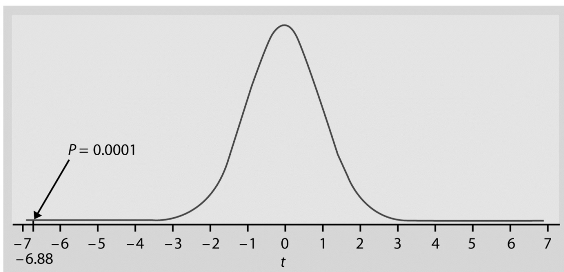
H_0: $\mu = 40$ vs. H_a: $\mu \neq 40$



Example (vit. C continued):

Test whether vit. C is lower than specifications.

H_0: $\mu = 40$ vs. H_a: $\mu < 40$



SAS

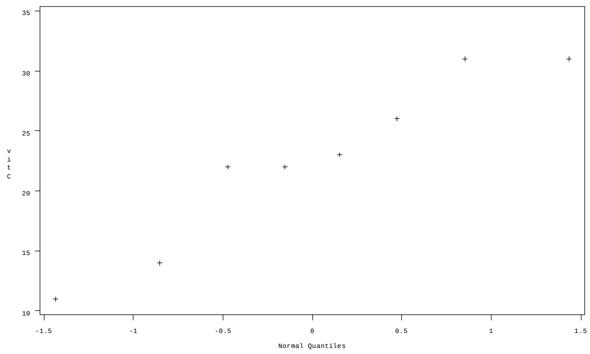
- **data** nowy;
- input vitC @@ ;
- datalines ;
- 26 31 23 22 11 22 14 31
- ;
- **run**;
- **proc univariate** data=nowy normal;
- qqplot;
- **run**;
- **proc ttest** h0=**40** alpha=**0.1**;
- var vitC;
- **run**;

Tests for normality

- Proc Univariate
- Tests for Normality

Test	--Statistic---		-----p Value------	
Shapiro-Wilk	W	0.918579	Pr < W	0.4184
Kolmogorov-Smirnov	D	0.222284	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.051067	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.322074	Pr > A-Sq	>0.2500

qqplot



TTest

- The TTEST Procedure
- Statistics

Var	N	Lower Mean	Mean	Upper Mean	Std Dev	Lower CL Std Dev	Upper CL Std Dev	Std Err
vitC	8	17.683	22.5	27.317	5.0728	7.1913	12.924	2.5425

- T-Tests

Variable	DF	t Value	Pr > |t|
vitC	7	-6.88	0.0002

Matched Pairs *t* Procedures

- Inference about a parameter of a single distribution is less common than comparative inference.
- In certain circumstances a comparative study makes use of single-sample ***t*** procedures.
- In a matched pairs study, subjects are matched in pairs; the outcomes are compared within each matched pair. *Compared=subtracted.*
- One typical situation here is "before" and "after" (quantitative) observations of the same subject.

Matched Pairs *t* Test

A matched pairs analysis is appropriate when there are *two measurements* or observations *per each individual* and we examine the change from the first to the second. Typically, the observations are "before" and "after" measures in some sense.

- For each individual, **subtract** the **"before"** measure **from** the **"after"** measure.
- **Analyze** the **difference** using the one-sample confidence interval or significance-testing ***t*** procedures (with H_0: μ=0).

Example

- **20** French teachers attend a course to improve their skills.
- The teachers take a Modern Language Association's listening test at the beginning of the course and at it's end.
- The maximum possible score on the test is **36**.
- The differences in each participant's "after" and "before" scores have sample mean **2.5** and sample standard deviation **2.893.**
- **Is the improvement significant?**
- **Construct a 95% confidence interval for the mean improvement (in the entire population).**

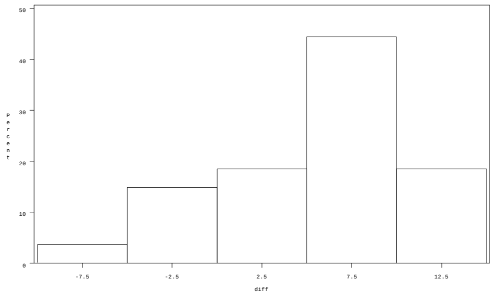
Example

- Loss of vitamin C content in storage and shipment
- Data – content of vit C in 18 bags of soy blend
- a) in the factory
- b) after 5 months and shipment to Haiti

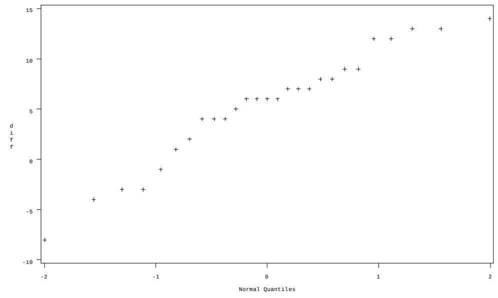
SAS

- **data** soy;
- infile 'c:/users/mbogdan/ECMI/data/ex07_039.txt' DLM='09'x;
- input id factory Haiti;
- **run**;
- **data** soy1;
- set soy;
- diff=factory-Haiti;
- **run**;
- **proc univariate** data=soy1 normal mu0=0;
- var diff;
- qqplot;
- histogram;
- **run**;

Histogram



QQplot



Tests for normality

- Tests for Normality
- Test --Statistic--- -----p Value------

Test	--Statistic---		-----p Value------	
Shapiro-Wilk	W	0.952279	Pr < W	0.2434
Kolmogorov-Smirnov	D	0.146454	Pr > D	0.1404
Cramer-von Mises	W-Sq	0.082695	Pr > W-Sq	0.1898
Anderson-Darling	A-Sq	0.482679	Pr > A-Sq	0.2197

Ttest (and nonparametric tests)

- Tests for Location: Mu0=0
- Test -Statistic- -----p Value------

Test	-Statistic-		-----p Value------	
Student's t	t	4.958945	Pr > |t|	<.0001
Sign	M	8.5	Pr >= |M|	0.0015
Signed Rank	S	152	Pr >= |S|	<.0001

Proc ttest

- **proc ttest** data=soy;
- paired factory*Haiti;
- **run**;

Results

- The TTEST Procedure
- Statistics

Difference	N	Lower CL Mean	Mean	Upper CL Mean	Lower CL Std Dev	Std Dev	Upper CL Std Dev	Std Err
factory - Haiti	27	3.1226	5.3333	7.5441	4.401	5.5884	7.6586	1.0755

- T-Tests

Difference	DF	t Value	Pr > |t|
factory - Haiti	26	4.96	<.0001

Robustness of *t* Procedures

- A statistical inference is called robust if it's outcome is not sensitive to the violation of the assumptions made.
- Real populations are never exactly normal.
- Usefulness of *t* procedures in practice depends on how strongly they are affected by non-normality.
- If they are not strongly affected, we say that they are robust.
- The *t* procedures are robust against non-normality of the population except in the case of outliers or strong skewness.

Robustness of *t* Procedures

- Practical guidelines for inference on a single mean:
  - Sample size < 15: Use *t* procedures if the data are close to normal; otherwise, don't.
  - Sample size ≥ 15: Use *t* procedures except in the presence of outliers or strong skewness.
  - Large sample size (≥ 40): Use *t* procedures even for clearly skewed distributions (but be careful with outliers).
  - Use normal quantile plot, histogram, stemplot or boxplot to investigate these properties of data.

Assumption that data are SRS--always important.

Nonparametric tests

- When the distribution is strongly different from normal
- Sign test
- N_+- number of observations for which
- var1 - var2 > 0
- Under H0 (distribution of var1-var2 is continous and symmetric around 0)
- N_+~

Sign test (ctd)

- N_-- number of observations for which
- var1 - var2 < 0
- Under H0 N_-~
- Test statistic
- $M=(N_+- N_-)/2$

Wilcoxon signed test

- Similar to the sign test but more powerful
- Method
  - Calculate var1-var2 in pairs
  - Assign ranks to absolute values of these differences (1 for the smallest, N for the largest)
  - Assign a sign for each rank (+ when Var1>var2, - when var1<var2)

- W_+ : sum of positive ranks
- $S= W_+ - N(N+1)/4$,
- Where N – number of observations for which var1 ≠ var 2

Obs	Y1	Y2	d	|d|	Rank	Signed Rank
1	33	25	8	8		
2	39	38	1	1		
3	25	27	-2	2		
4	29	20	9	9		
5	50	54	-4	4		
6	45	40	5	5		
7	36	30	6	6		

- Wilcoxon signed test is more powerful then the sign test.
- Sign test can be used when the data are coded in terms of preferences rather than numbers (e.g. better/worse, yes/no, smaller/larger)

Comparing Two Independent Samples

- The two-sample ***z*** statistics
- The two-sample ***t*** procedures:
  - – significance test
  - – confidence interval
- Robustness of the two-sample procedures
- Small samples

Two-Sample Problems

- Two-sample problems typically arise from a randomized comparative experiment with two treatment groups. (Experimental study.)
- Comparing random samples separately selected from two populations is also a two-sample problem. (Observational study.)
- Unlike matched pairs design, there is no matching of the units in the two sample, and the two samples may be of different sizes.

Notation for Two-Sample Settings

Population	Population mean	Population standard deviation
1	μ_1	σ_1
2	μ_2	σ_2

Notation for Two-Sample Settings

- Suppose an SRS of size n_1 is selected from the 1st population, and another SRS of size n_2 is selected from the 2nd population.

Population	Sample size	Sample mean	Sample standard deviation
1	n_1	$\bar{x}_1$	s_1
2	n_2	$\bar{x}_2$	s_2

Example (metabolism rates for men and women):

Obs	Gender	Mass	Rate
1	M	62	1792
2	M	62.9	1666
3	F	36.1	995
4	F	54.6	1425
5	F	48.5	1396
6	F	42	1418
7	M	47.4	1362
8	F	50.6	1502
9	F	42	1256
10	M	48.7	1614
11	F	40.3	1189
12	F	33.1	913
13	M	51.9	1460
14	F	42.4	1124
15	F	34.5	1052
16	F	51.1	1347
17	F	41.2	1204
18	M	51.9	1867
19	M	46.9	1439

- **data** metabolism;
- infile 'c:/users/mbogdan/ECMI/data/metabolism.txt';
- input id gender $ mass rate;
- **run**;
- **proc sort** data=metabolism out=met2;
- by gender;
- **run**;
- **PROC BOXPLOT** data=met2 ;
- PLOT rate*gender;
- **run**;

The Two-Sample *z* Statistic

- A natural estimator of the difference $\mu_1 - \mu_2$ is the difference between the sample means
- From the rules of adding means and variances: D= $\bar{x}_1 - \bar{x}_2$

  (population) mean of differences: $\mu_1 - \mu_2$

  (population) SD of differences of sample standard deviations: $\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$
- This expresses the mean and variance of the distribution of *differences* (of sample means) in terms of the parameters of the two original populations.
- If the two population distributions are both normal, then the distribution of D is also normal.

About the distribution of the Two-Sample *z* Statistic

- Suppose that $\bar{x}_1$ is the mean of an SRS of size n_1 drawn from an N(μ_1, σ_1) population and that $\bar{x}_2$ is the mean of an independent SRS of size n_2 drawn from an N(μ_2, σ_2) population. Then the two-sample *z* statistic

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

has the standard normal *N*(0, 1) as its sampling distribution.

Inference – Two Populations, Known Population Standard Deviations

- If μ_1 and μ_2 are unknown, then a level C confidence interval for $\mu_1 - \mu_2$ is

$$\bar{x}_1 - \bar{x}_2 \pm z^* \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

where $P(-z^* \leq Z \leq z^*) = C$.

Inference – Two Populations, Known Population Standard Deviations

- We want to test H_0: $\mu_1 = \mu_2$ against one of the following alternative hypotheses:
  - H_a: $\mu_1 > \mu_2$
  - H_a: $\mu_1 < \mu_2$
  - H_a: $\mu_1 \neq \mu_2$
- The **z** test statistic when known population SDs:

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

Inference – Two Populations, Known Population Standard Deviations

Alternative Hypothesis	P-value
H_a: $\mu_1 > \mu_2$	$P(Z>z)$
H_a: $\mu_1 < \mu_2$	$P(Z<z)$
H_a: $\mu_1 \neq \mu_2$	$2*P(Z>\mid z \mid)$

Facts about distribution– Two Populations, **Unknown** Population Standard Deviations

- Suppose μ_1, μ_2, σ_1 and σ_2 are unknown.
- Two-sample ***t*** statistic for difference in means:

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

- is *approximately* ***t*** distributed with df either approximated by software or taken as:

  min(n_1 – 1, n_2 – 1).

Inference – Two Populations, Unknown Population Standard Deviations

- If μ_1, μ_2, σ_1, and σ_2 are unknown, then a level C confidence interval for $\mu_1 - \mu_2$ is

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

where $P(t^*_{df} \le t \le t^*_{df}) = C$

So, this is ***t***. Degrees of freedom as before: min(n_1 – 1, n_2 – 1) or from software.

TABLE D *t* distribution critical values

	Upper tail probability *p*											
df	.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	1.638	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
8	0.706	0.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	0.703	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	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878	3.197	3.611	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861	3.174	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845	3.153	3.552	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831	3.135	3.527	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819	3.119	3.505	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807	3.104	3.485	3.768
24	0.685	0.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797	3.091	3.467	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787	3.078	3.450	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.162	2.479	2.779	3.067	3.435	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.158	2.473	2.771	3.057	3.421	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.154	2.467	2.763	3.047	3.408	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.150	2.462	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	3.174	3.390
1000	0.675	0.842	1.037	1.282	1.646	1.962	2.056	2.330	2.581	2.813	3.098	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*											

Example (*Metabolism rate. Women vs. men.)*

$$n_1 = 12,\ \ \bar{x}_1 = 1235.1,\ \ s_1 = 188.3$$
$$n_2 = 7,\ \ \bar{x}_2 = 1600,\ \ \ s_1 = 189.2$$

Find difference in mean metabolism rates between men and women.

Solution:

Inference – Two Populations, Unknown Population Standard Deviations

- Goal: test H_0: $\mu_1 = \mu_2$ against one of the following alternative hypotheses when σ_1, σ_2 are unknown:
  - – H_a: $\mu_1 > \mu_2$
  - – H_a: $\mu_1 < \mu_2$
  - – H_a: $\mu_1 \neq \mu_2$
- The ***t*** test statistic is

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

- Degrees of freedom as before: min(n_1 – 1, n_2 – 1) of from software.

Inference – Two Populations, Unknown Population Standard Deviations σ_1, σ_2

Alternative Hypothesis	P-value
H_a: $\mu_1 > \mu_2$	$P(T \geq t)$
H_a: $\mu_1 < \mu_2$	$P(T \leq t)$
H_a: $\mu_1 \neq \mu_2$	$2*P(T \geq \mid t \mid)$

Example: Do women have different metabolism rate than men?

Robustness of Two-Sample *t* Test

- The two-sample *t* procedures are even more robust than the one-sample *t* methods. They are robust against non-normal population distributions, in particular if the population distributions are symmetric and the two sample sizes are equal.
- Outliers are always a problem: may need to be eliminated. Skewness less important for not-too-small sample sizes.
- t procedures are rather conservative so your calculated P-values may be even larger than in reality. This is good (safe).

Degrees of freedom for two-sample t procedures

- df as before: min(n_1 – 1, n_2 – 1) of from software
- The choice of min(n_1 – 1, n_2 – 1) is conservative.
- Software will usually give smaller P-values.
- In our example with metabolism rates software (calculator☺) will give df=12.6
- Here no difference in final conclusion…

SAS

- **proc ttest** data=metabolism ci=equal;
- class gender;
- var rate;
- **run**;

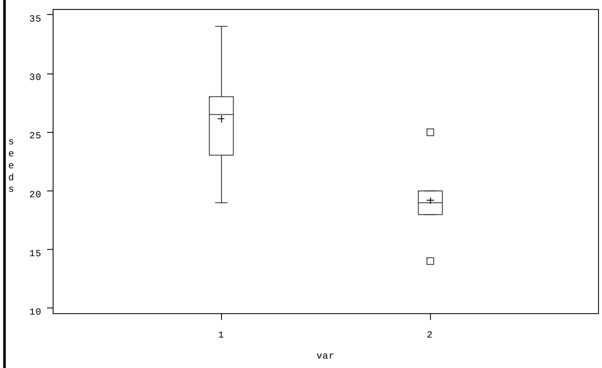
Wilcoxon-Mann-Whitney test

- Distribution strongly different from normal (outliers, strong skewness)
- Method
- Aggregate the data from both samples
- Assign rank for each observation (comparing to all observations in both groups)
- Compute the sum of ranks in both groups, R1 and R2
- Test statistics W=min(R1, R2)

Example

- Number of seeds produced by two varieties of a certain plant.
- Data:
- Var 1: 19, 23, 25, 28, 28, 34 (n1 = 6)
- Var 2: 14, 18, 19, 20, 25 (n2 = 5)

- **proc boxplot** data=seeds;
- plot seeds*var/ boxstyle=schematic;
- **run**;
- **proc npar1way** data=seeds wilcoxon;
- class var;
- var seeds;
- exact wilcoxon;
- **run**;



Wilcoxon Scores (Rank Sums) for Variable seeds
Classified by Variable var

var	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
1	6	47.0	36.0	5.439753	7.833333
2	5	19.0	30.0	5.439753	3.800000

- Exact Test

- One-Sided Pr <= S 0.0238
- Two-Sided Pr >= |S - Mean| 0.0498