Comparing groups

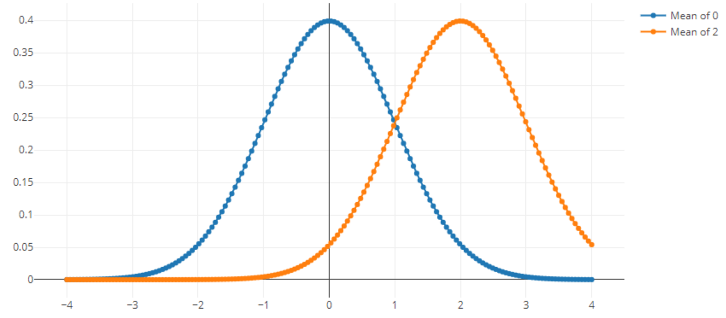
Chi-square test

T-test

E0420

Week 3

Handed-ness / Sex	Right-handed	Left-handed	Total
Male	43	9	52
Female	44	4	48
Total	87	13	100



Let's analyze!

- https://stats.idre.ucla.edu/other/mult-pkg/whatstat/
- Testing associations between:
- 1 categorical independent variable (IV, predictor, exposure variable)
  - T-test can handle only 2 categories (if more, use ANOVA)
  - Chi-square is not limited by the number of categories
- 1 dependent variable (DV, outcome)
  - Categorical – Chi-square test
  - Continuous – T-test

Let's analyze!

- Categorical IV:
  - Gender (males vs females), type of exposure (eating fish vs not eating fish), work or school group/class (scientists vs administrators), or experimental/treatment group
- DV
  - Continuous – BMI, depression score, hrs slept per night
  - Categorical – presence of a diagnosis (diabetes, pregnancy, depression), success or failiure
- Comparing means (T-test) or proportions (Chi-square) of a DV across 2 groups/levels of an IV

- A word about categorization of continuous IVs
  - Mean/median/tercile split
  - Change in the original information, smaller effect size, potentially spurious effects

How does Chi-square work?

1. Constructing contingency table (2x2 table, cross tabulation)

= number of cases in each category

2. Comparing observed numbers in each category to expected numbers in each category if there is no association (= null hypothesis)

= obtaining the $\chi 2$ statistic

3. Using $\chi 2$ distribution for specific degrees of freedom to determine how likely is the obtained $\chi 2$ if null hypothesis is true

= determining statistical significance (*p*-value)

- If the probability is 5% ($\alpha = .05$) or less, the $\chi 2$ is stat. sig.
- In general, large $\chi 2$ suggests rejection of null hypothesis

1. Contingency table

	Feeling depressed		
	Yes	No	Total
Male	96 (40%)	144 (60%)	240
Female	72 (24%)	228 (76%)	300
Total	168 (31%)	372 (69%)	540

Observed numbers

2. Obtaining the $\chi 2$ statistic

Observed numbers

	Feeling depressed		
	Yes	No	Total
Male	96 (40%)	144 (60%)	240
Female	72 (24%)	228 (76%)	300
Total	168 (31%)	372 (69%)	540

Expected numbers

	Feeling depressed		
	Yes	No	Total
Male	**75**	**165**	240
Female	**93**	**207**	300
Total	168 (31%)	372 (69%)	540

2. Obtaining the χ2 statistic

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

χ^2 = chi squared

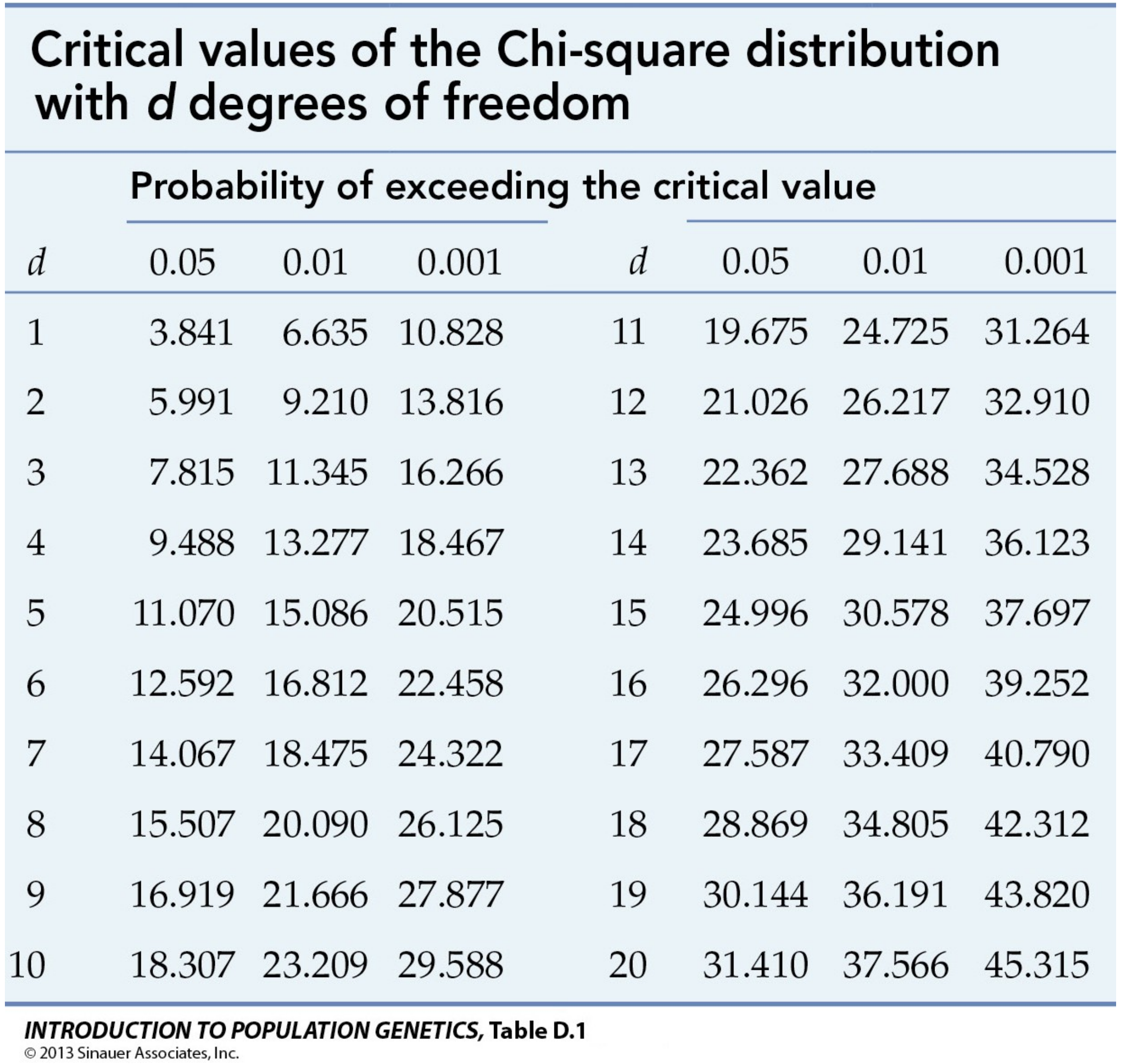
O_i = observed value

E_i = expected value

- In our case, χ2 = 15.97

3. Determining statistical significance

- **degrees of freedom (df) = (r-1)(c-1)** where r is the number of rows and c is the number of columns
- Critical value is based on the selected alpha level and df
- In our case, $\chi 2 = 15.97$



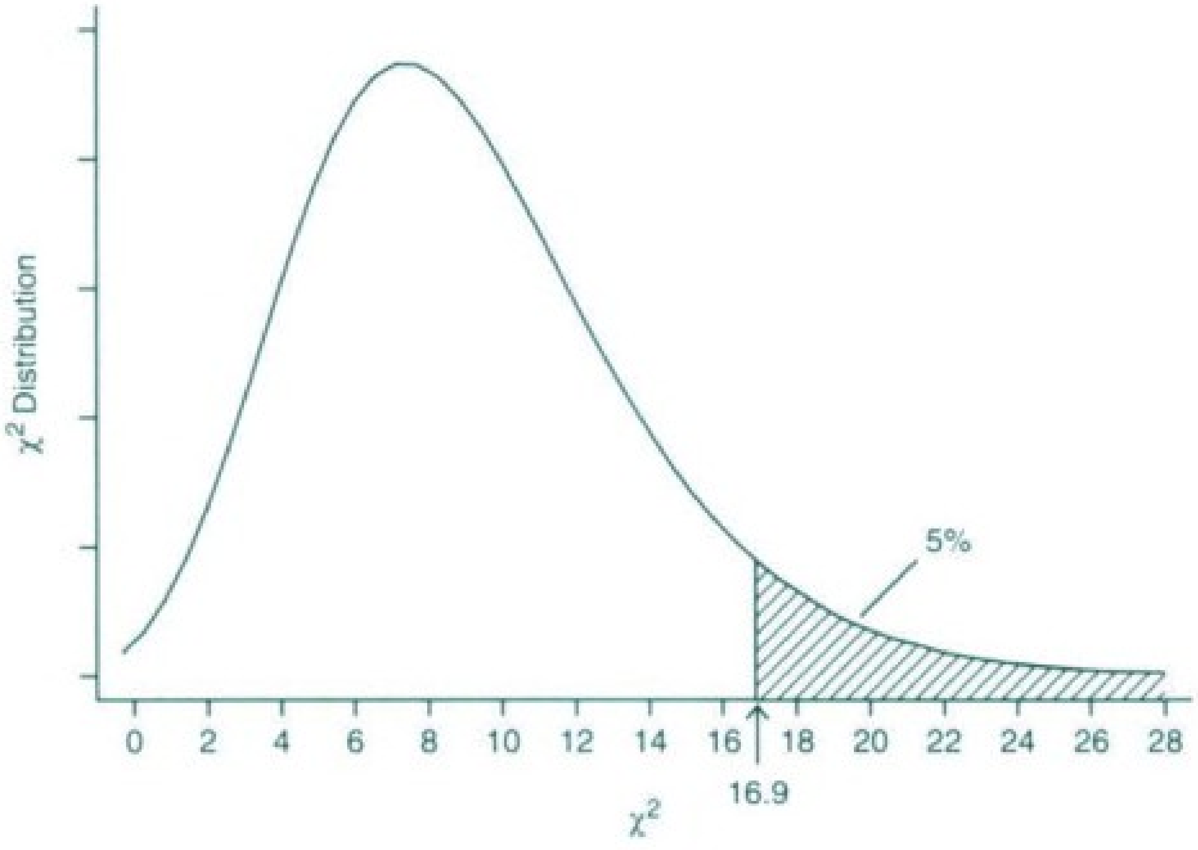
Critical values of the Chi-square distribution with *d* degrees of freedom

	Probability of exceeding the critical value						
d	0.05	0.01	0.001	*d*	0.05	0.01	0.001
1	3.841	6.635	10.828	11	19.675	24.725	31.264
2	5.991	9.210	13.816	12	21.026	26.217	32.910
3	7.815	11.345	16.266	13	22.362	27.688	34.528
4	9.488	13.277	18.467	14	23.685	29.141	36.123
5	11.070	15.086	20.515	15	24.996	30.578	37.697
6	12.592	16.812	22.458	16	26.296	32.000	39.252
7	14.067	18.475	24.322	17	27.587	33.409	40.790
8	15.507	20.090	26.125	18	28.869	34.805	42.312
9	16.919	21.666	27.877	19	30.144	36.191	43.820
10	18.307	23.209	29.588	20	31.410	37.566	45.315

***INTRODUCTION TO POPULATION GENETICS,* Table D.1**

https://www.mun.ca/biology/scarr/4250_Chi-square_critical_values.html

Chi-square distribution



Chi-square write-up

- A chi-square test was performed to examine the association between gender and feeling depressed. The association between these variables was significant, $\chi 2$ (1, *N* = 540) = 15.97, $p < .001$. Women were more likely than men to feel depressed.

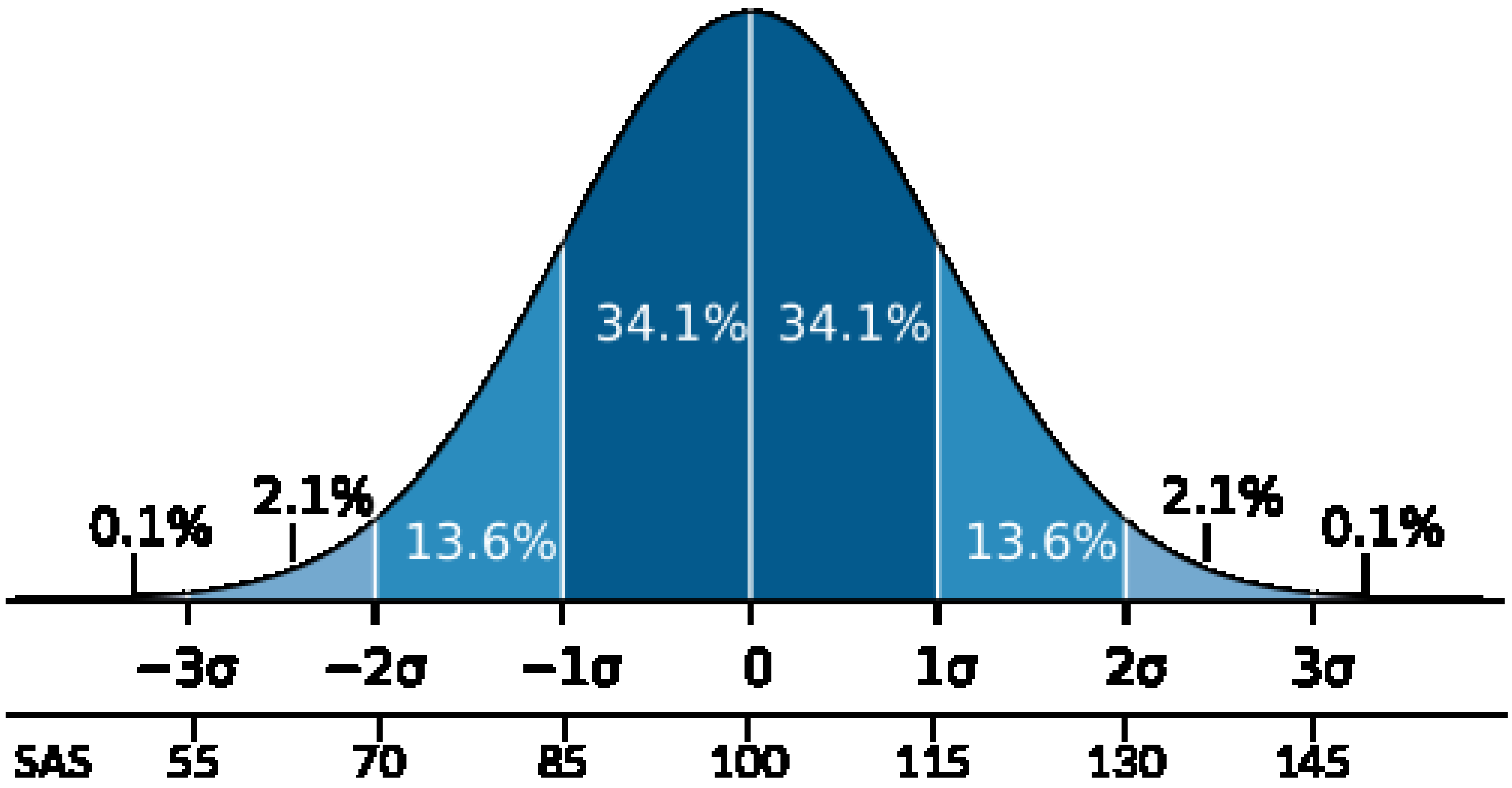
T-test use

- Purpose:
  - Comparing means of a continuous DV across 2 groups (binary IV)
  - Useful when we have „natural“ IV groups (e.g., experimental condition)
- Types:
  - Independent T-test
  - Paired-samples T-test

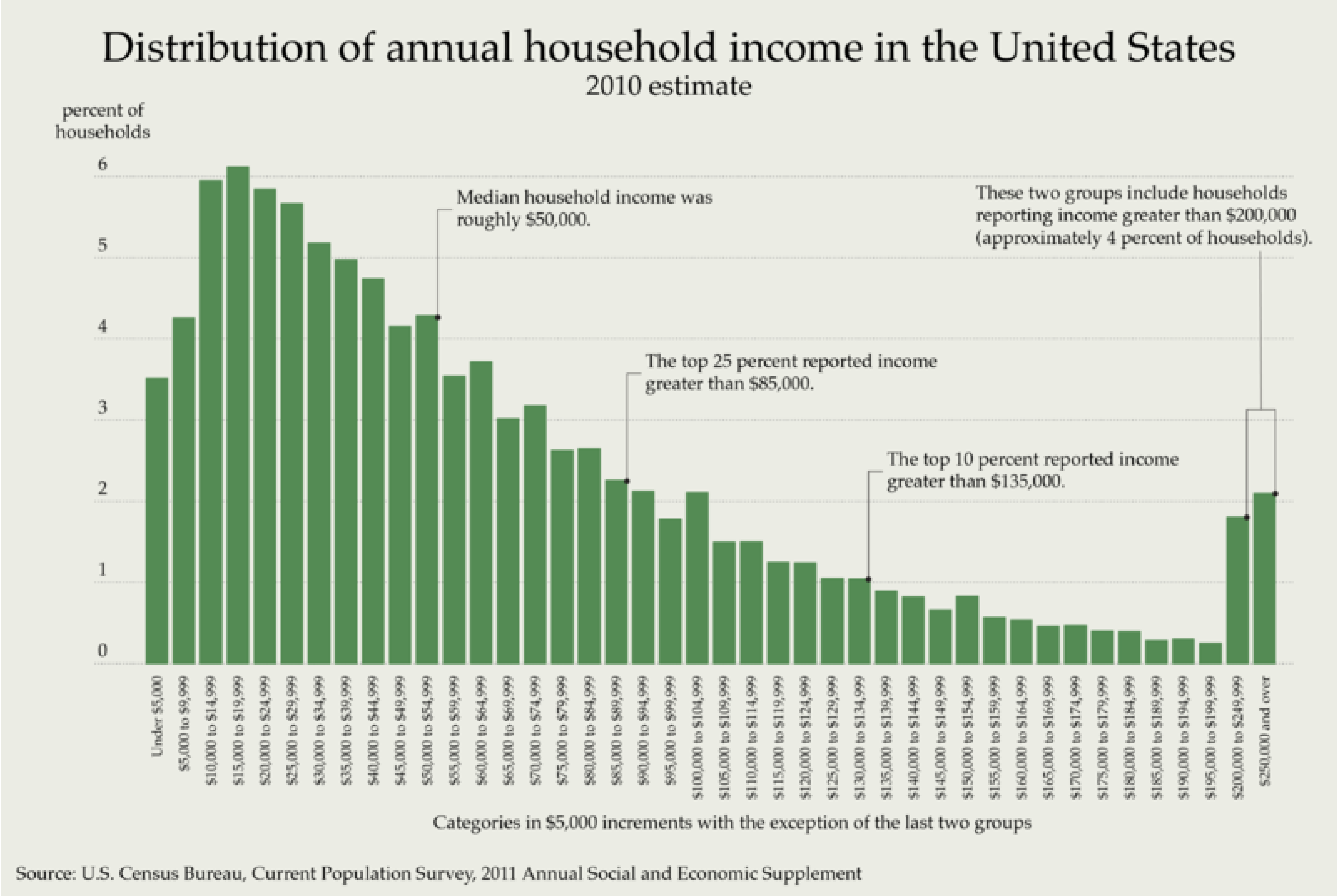
T-test assumptions

- Assumption of independence
  - Applies for independent t-test
- Assumption of normality
  - DV should be approximately normally distributed
- Assumption of homogeneity of variance
  - The two independent samples are assumed to be drawn from populations with identical population variances
  - The variances of DV should be equal in both IV groups

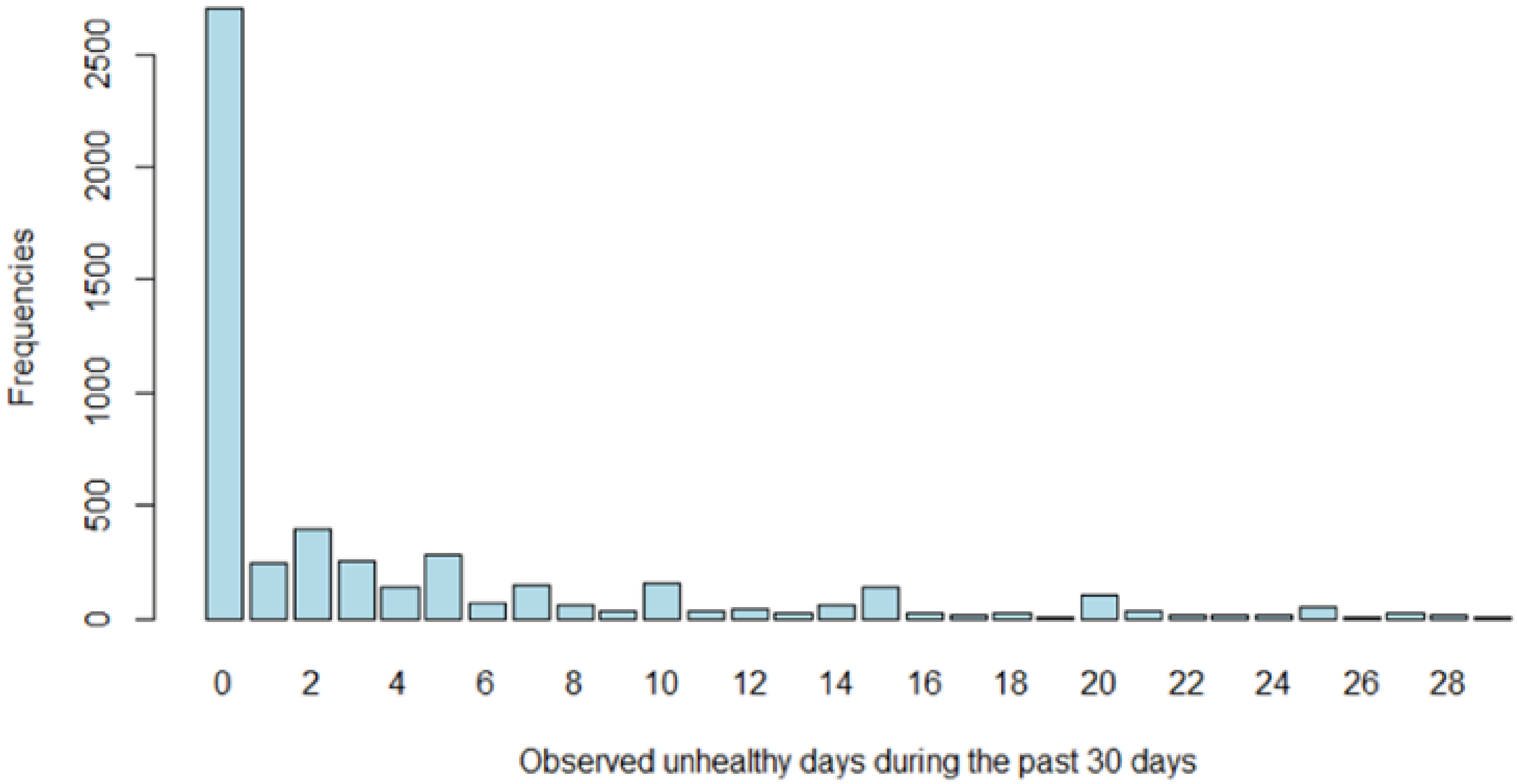
Normal/Gaussian distribution



https://commons.wikimedia.org/wiki/File:Wechsler.svg



https://commons.wikimedia.org/wiki/File:Distribution_of_Annual_Household_Income_in_the_United_States_2010.png



Frequencies
0
500
1000
1500
2000
2500
0
2
4
6
8
10
12
14
16
18
20
22
24
26
28
Observed unhealthy days during the past 30 days

How does T-test work?

1. Obtaining mean and SD of the DV in each IV group

2. Comparing the observed differences in the group means to expected differences (= null hypothesis)

= obtaining the *t* statistic

3. Using *t* distribution for specific degrees of freedom to determine how likely is the obtained *t* if null hypothesis is true

= determining statistical significance (*p*-value)

- If the probability is 5% (α = .05) or less, the *t* is stat. sig.
- In general, large *t* suggests rejection of null hypothesis

1. Obtaining means and SDs

Testing differences in depression scores on Beck Depression Inventory (BDI) between males and females

- Females
  - Mean = 9
  - SD = 2
  - N = 50
- Males
  - Mean = 6
  - SD = 3
  - N = 40

2. Obtaining the *t* statistic

Paired-samples T-test

$$t = \frac{\bar{x}_{\text{diff}} - 0}{s_{\bar{x}}}$$

where

$$s_{\bar{x}} = \frac{s_{\text{diff}}}{\sqrt{n}}$$

where

$\bar{x}_{\text{diff}}$ = Sample mean of the differences

n = Sample size (i.e., number of observations)

s_{diff} = Sample standard deviation of the differences

$s_{\bar{x}}$ = Estimated standard error of the mean (s/sqrt(n))

Independent T-test

Equal variance assumed

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

with

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

Where

$\bar{x}_1$ = Mean of first sample

$\bar{x}_2$ = Mean of second sample

n_1 = Sample size (i.e., number of observations) of first sample

n_2 = Sample size (i.e., number of observations) of second sample

s_1 = Standard deviation of first sample

s_2 = Standard deviation of second sample

s_p = Pooled standard deviation

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

Equal variance not assumed

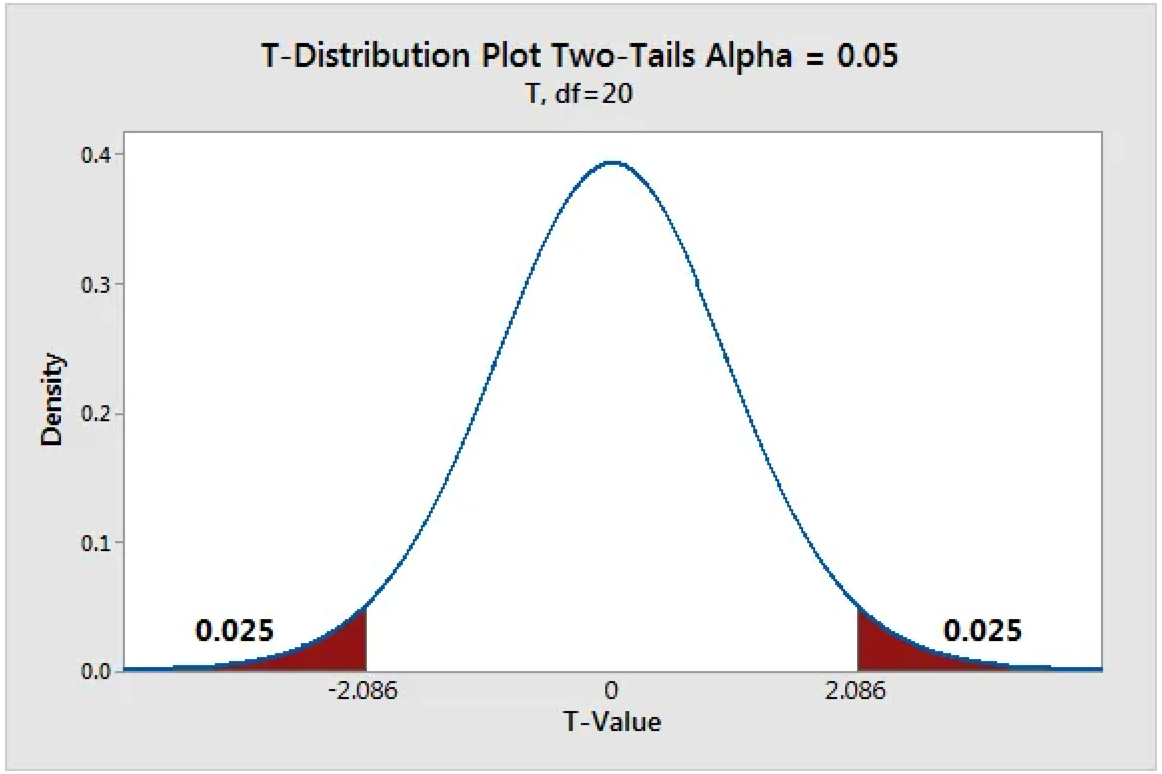
3. Determining statistical significance

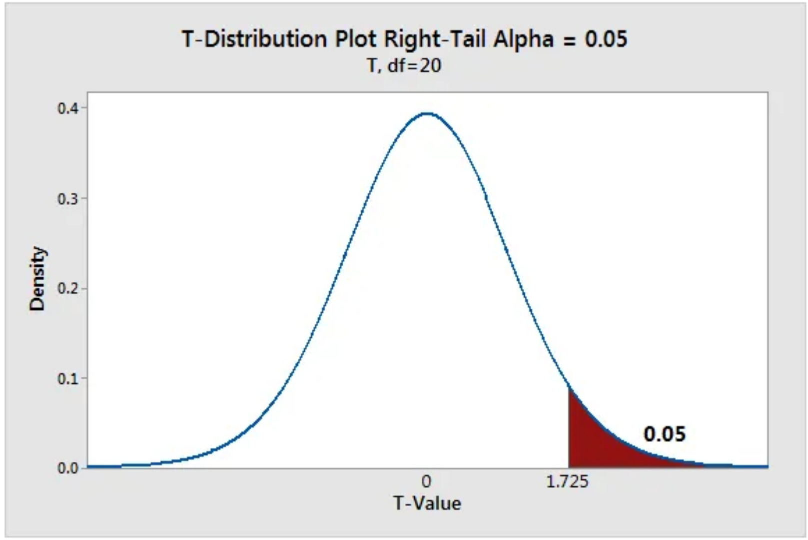
- df = *n - 1* for paired-samples
- df = $(n_1 - 1) + (n_2 - 1)$ for independent samples
- Critical value is based on the selected alpha level, df, and whether we are testing one-tailed or two-tailed hypothesis

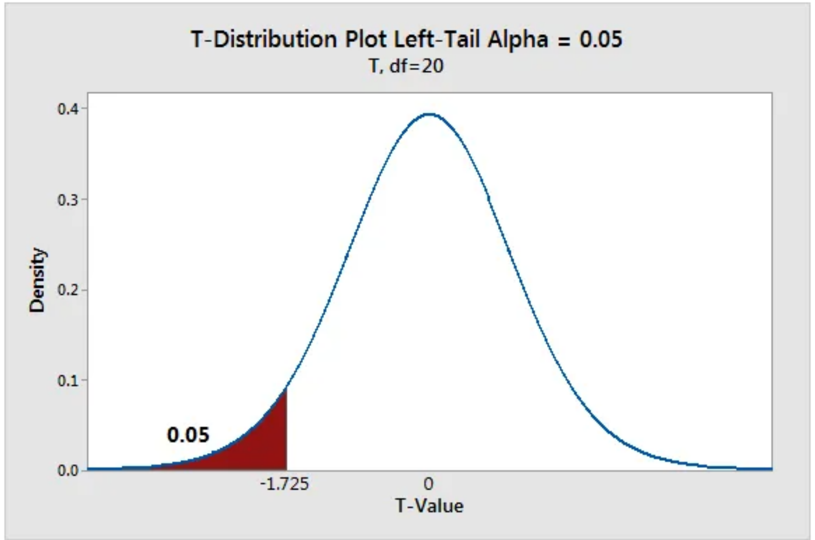
- In our case, $t = 5.67$

	P						
one-tail	0.1	0.05	0.025	0.01	0.005	0.001	0.0005
two-tails	0.2	0.1	0.05	0.02	0.01	0.002	0.001
DF							
1	3.078	6.314	12.706	31.821	63.656	318.289	636.578
2	1.886	2.92	4.303	6.965	9.925	22.328	31.6
3	1.638	2.353	3.182	4.541	5.841	10.214	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.61
5	1.476	2.015	2.571	3.365	4.032	5.894	6.869
6	1.44	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.86	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.25	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.93	4.318
13	1.35	1.771	2.16	2.65	3.012	3.852	4.221
14	1.345	1.761	2.145	2.624	2.977	3.787	4.14
15	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	1.337	1.746	2.12	2.583	2.921	3.686	4.015
17	1.333	1.74	2.11	2.567	2.898	3.646	3.965
18	1.33	1.734	2.101	2.552	2.878	3.61	3.922
19	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	1.325	1.725	2.086	2.528	2.845	3.552	3.85
21	1.323	1.721	2.08	2.518	2.831	3.527	3.819
22	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	1.319	1.714	2.069	2.5	2.807	3.485	3.768
24	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	1.316	1.708	2.06	2.485	2.787	3.45	3.725
26	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	1.314	1.703	2.052	2.473	2.771	3.421	3.689
28	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	1.311	1.699	2.045	2.462	2.756	3.396	3.66
30	1.31	1.697	2.042	2.457	2.75	3.385	3.646
60	1.296	1.671	2	2.39	2.66	3.232	3.46
120	1.289	1.658	1.98	2.358	2.617	3.16	3.373
1000	1.282	1.646	1.962	2.33	2.581	3.098	3.3
Inf	1.282	1.645	1.96	2.326	2.576	3.091	3.291

One-tailed or two-tailed hypothesis







T-test write-up

- The 25 participants who received the drug intervention (*M* = 480, *SD* = 34.5) compared to the 28 participants in the control group (*M* = 425, *SD* = 31) demonstrated significantly better peak flow scores, $t(51) = 2.1$, $p = .04$.
- There was a significant increase in the volume of alcohol consumed in the week after the end of semester (*M* = 8.7, *SD* = 3.1) compared to the week before the end of semester (*M* = 3.2, *SD* = 1.5), $t(52) = 4.8$, $p < .001$.