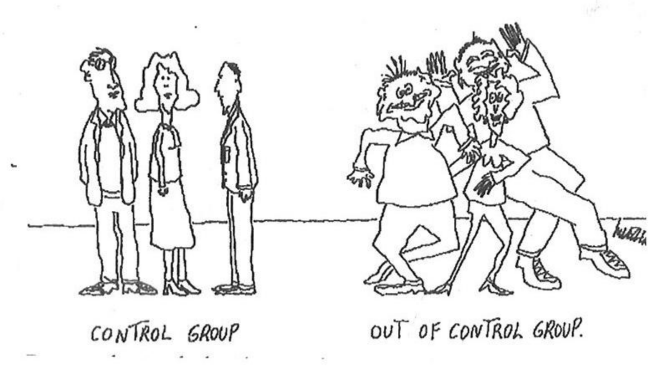
Hypothesis Testing:
Two Samples



Comparing Two Samples:
Testing for Mean Differences

- Common research question: Are two groups different?
- We want to compare means of two groups
- For example: by gender, race/ethnicity, social class, etc.

Comparing Two Samples:
Testing for Mean Differences

- Or: experimental group vs control group
- Less common in sociology but common in biomedical research – mice…

Or humans…

"And this is the control group which we keep on 200 cl of decaf latte."

Hypotheses for Two Means

- Null hypothesis: The means of two groups are the same in the population
- Research hypothesis: The means of two groups are different in the population (if directional: mean 1 is larger mean 2, or vice versa)

Example from Hypothesis Writing Exercise

- Question: Do Whites and African Americans differ in the amount of time they spend with their relatives?
- H0: Whites and African Americans are similar in the amount of time they spend with relatives
- H1 (non-directional): Whites and African Americans differ in the amount of time they spend with relatives
- H1 (directional): Whites spend less time with relatives than African Americans

Comparing Two Samples: Step-by-Step

1. State hypotheses:

- H0: $\mu_1 = \mu_2$
- H1: $\mu_1 < \mu_2$
- H1: $\mu_1 > \mu_2$

directional

- H1: $\mu_1 \neq \mu_2$

non-directional

2. Select alpha: 0.05, 0.01, .001, .10

3. Test statistic: Student's t (which is very similar to z scores after sample size reaches 30)

Comparing Two Samples: Step-by-Step

4. Formula:

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\left[\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}\right]\left[\frac{n_1+n_2}{n_1 n_2}\right]}}$$

$\overline{X}_1$	is the mean for Group 1
$\overline{X}_2$	is the mean for Group 2
n_1	is the number of participants in Group 1
n_2	is the number of participants in Group 2
s_1^2	is the variance for Group 1
s_2^2	is the variance for Group 2

5. Use the table to find critical value: Table B2 (df=$n_1 + n_2$ -2, alpha, one-tailed vs two-tailed)

6. Compare computed value and critical value

7. State your decision about H0

Example: Comparing Two Groups

- Want to study the effectiveness of a children's TV program designed to teach reading skills
- 50 children will participate in the study: 25 are randomly assigned to the experimental group, 25 – to control group. Over a few months, the experimental group watches the "reading" show, the control group watches another show
- After that, the mean reading score in the experimental group = 12.3 (SD = 2.1), and in the control group – 11.2 (SD=1.8)
- Can we conclude with 95% certainty that our program helps children learn how to read?

Example: Comparing Two Groups, Step by Step

1. H_0: $\mu_1 = \mu_2$

H_1: $\mu_1 > \mu_2$-- since we expect higher scores for the experimental group→ use a directional research hypothesis → one-tailed test

2. We want to be 95% confident if we conclude that this program works – we need to use significance level alpha = 0.05 (1-.95=.05).

3. Test statistic – Student's t

Example: Comparing Two Groups, Step by Step

$$t = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{\left[(n_1-1)s_1^2 + (n_2-1)s_2^2\right]\ (n_1+n_2)}{(n_1+n_2-2)(n_1n_2)}}}$$

4. Compute:

Numerator: 12.3-11.2=1.1

Denominator:

sqrt[((25-1)*2.1*2.1 + (25-1)*1.8*1.8)/(25+25-2)* (25+25)/25*25] = 0.55

t= 1.1 / 0.55 = 2

5. Use Table B2 to find the critical value: df=n_1+n_2-2=50-2=48, alpha=0.05, one-tailed test → t_{crit} = 1.676

6. Does computed statistic exceed critical value? 2 is larger than 1.676 so it does exceed the critical value

7. Conclusion: We can reject the null hypothesis H_0: $\mu_1 = \mu_2$

8. Those who watch the reading program are doing significantly better than those who watch another show → our program likely works

Chances of Type I and Type II Errors

- Reject the null:
  - By hand: chances of type I error < alpha
  - In Stata: chances of type I error <p-value
  - Type II error probability = 0
- Fail to reject the null:
  - Type I error probability = 0
  - Large sample → chances of type II error are small
  - Small sample → chances of type II error are large

Chances of Error?

- Alpha=.05 → chances of Type I error are <5%
- We rejected H0 → chance of Type II error = 0

"Ed is in a study involving diabetes and lack of exercise. This is his remote control group."

© 2006 Diabetes Health

Two Independent Samples in Stata: Stating the Problem

- Are self-employed individuals older on average than those employed by others?
- H0: Self-employed and employees are the same age on average in the U.S. population.
- H1: On average, self-employed are older than employees in the U.S. population.

H_0: $\mu_1 = \mu_2$

H_1: $\mu_1 > \mu_2$

Hypotheses Testing about Means in Stata: Two Independent Samples

`ttest age, by(wrkslf)`

```
Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
SELF-EMP |     194    54.78351    1.129589    15.73335    52.55558    57.01143
 SOMEONE |    1696    48.00825    .4220768    17.38218    47.18041     48.8361
---------+--------------------------------------------------------------------
combined |    1890     48.7037    .3988291    17.33875    47.92151     49.4859
---------+--------------------------------------------------------------------
    diff |             6.77525    1.305186                4.215492    9.335009
------------------------------------------------------------------------------
    diff = mean(SELF-EMP) - mean(SOMEONE)                         t =   5.1910
Ho: diff = 0                                     degrees of freedom =     1888

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 1.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 0.0000
```

`t=5.1910, p<.001 → reject the null`

Comparison: P-Value and Alpha or Test Statistic and Critical Value

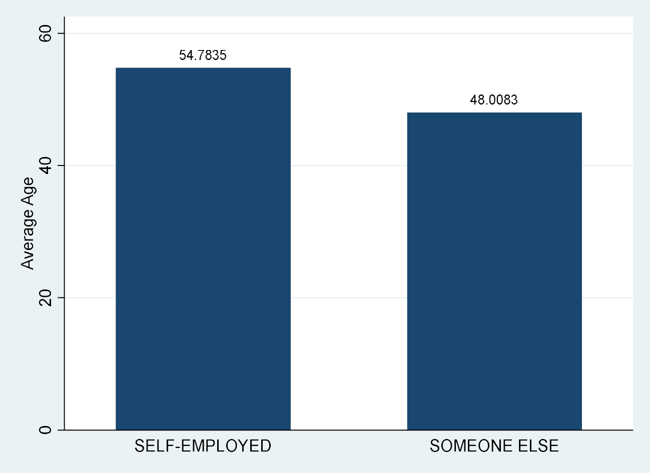
- Compare p-value to alpha:
  - p-value smaller than alpha → reject H0
  - P-value larger or equal to alpha → fail to reject H0
- When using Student's t, can't use the table to calculate p-values
- Compare test statistic (e.g., z score we calculated above) to a critical value (e.g., value of z corresponding to our chosen alpha)
  - Test statistic larger than critical value → reject H0
  - Test statistic smaller than critical value → fail to reject H0

Conclusion

- We are 99.9% confident that among Americans, self-employed individuals are older on average than those employed by others.
- Chances of Error:
  - P-value <.001 → chances of Type I error are <0.1% (probability <.001).
  - We rejected H0 → chance of Type II error = 0

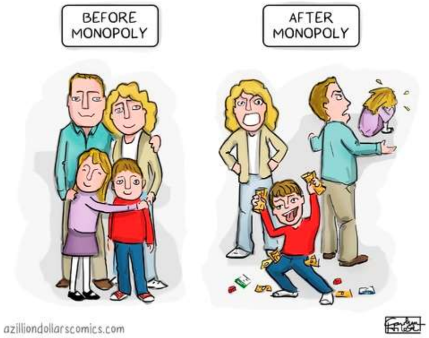
Graphs for Two Groups: Bivariate Bar Graph

```
graph bar age, over(wrkslf) blabel(bar) ytitle("Average Age")
```



Caution: Two Independent Groups vs Paired Samples

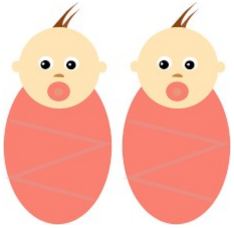
- Independent samples: No meaningful pairs (can sort in any way)
- The most typical paired samples case: Pre-test and post-test (before and after measures)



Other Common Paired Samples

- Husbands and wives (rather than unrelated women and men), children and parents, pairs of siblings
- Different measures for the same person (e.g., weights lifted with right arm vs left arm; reading score vs math score)

A statistician's wife had twins. He was delighted. He rang the minister who was also delighted. “Bring them to church on Sunday and we'll baptize them,” said the minister. “No,” replied the statistician. “Baptize one. We'll keep the other as a control.”



Paired Data: Focus on Differences within Pairs

- Can’t look at two means and SD separately because we need to deal with pairs!
- We need to calculate mean difference and the standard deviation of the difference
- Collapsing from two variables into one, and then it’s a single mean test of H0: mean=0
- As a result, we use:

$$t = \frac{\overline{d}}{S_d / \sqrt{n}}$$

(df= number of pairs – 1)

Example: Paired T-Test

- We want to study the effectiveness of a financial literacy video that we developed. For 15 volunteers, we measure their financial literacy before they watch the video and then again one week after they watch the video. Can we conclude with 95% certainty that this video increases financial literacy?

Person #	Before	After
1	12	12
2	13	15
3	11	12
4	10	9
5	13	16
6	11	13
7	10	13
8	9	11
9	9	12
10	10	10
11	12	15
12	11	9
13	11	14
14	14	16
15	11	16

- Mean of D = 1.73
- SD of D = 1.83
- T = 1.73/(1.83/sqrt(15)) = 3.66
- Df=14, alpha=.05
- Two-tailed
- T critical = 2.145
- Reject H0

Person #	Before	After	D
1	12	12	0
2	13	15	2
3	11	12	1
4	10	9	-1
5	13	16	3
6	11	13	2
7	10	13	3
8	9	11	2
9	9	12	3
10	10	10	0
11	12	15	3
12	11	9	-2
13	11	14	3
14	14	16	2
15	11	16	5

Conclusion and Error Chances

- Conclusion: We are 95% confident that our video helps increase levels of financial literacy

Error Chances:

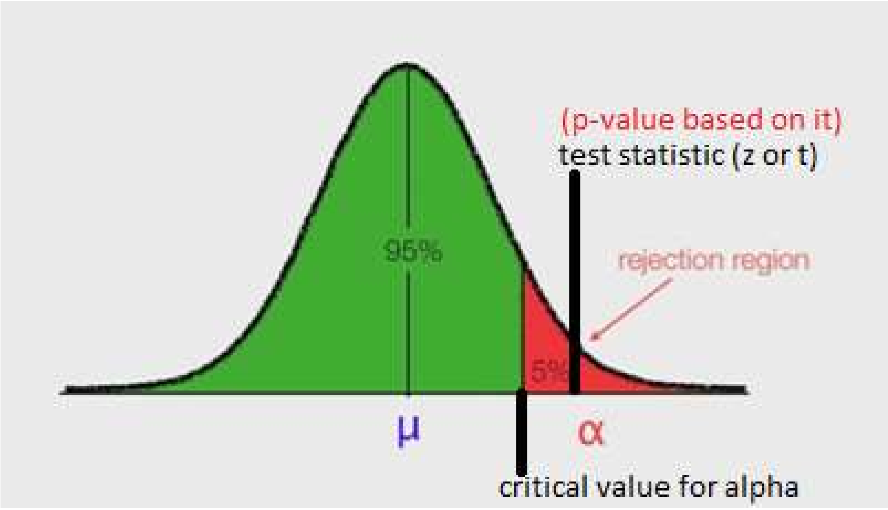
- Alpha=.05 → Chances of Type I error <5%
- Chances of Type II error = 0 after we reject H0

The Meaning of Alpha and P-Value

- By hand: we reject H0 when computed statistic > critical value
- Critical value cuts off the critical region defined by our alpha
- Alpha = .05 → computed statistic will be in the critical region by chance in < 5% of cases
- That means, risk of Type I error < 5%

So What About P-Value?

- In Stata: we reject H0 when p-value is below alpha
- P-value = exact percentage of cases when the computed statistic will be as extreme or more extreme solely by chance (assuming the null is correct)
- P-value gives us a more exact estimate of probability of Type I error
- If p=.019, the chance of Type I error is <1.9% (so we are 98.1% confident in rejecting null)
- More exact than just using alpha=.05 and assuming 5% chance of Type I error



- P-value calculator: https://www.socscistatistics.com/pvalues/tdistribution.aspx

Confidence

- Confidence level + probability of Type I error (alpha) = 1
- Power of statistical test + probability of Type II error (beta) = 1
- If we reject null, we can say the exact level of confidence based on alpha (or p-value): e.g., we say “99% confident” if p<.01 (1% chance of Type I error, 0% chance of Type II)
- If we fail to reject the null, the confidence level based on our alpha <u>does not apply</u> (0% chance of Type I error but probability of Type II error is usually higher than .05)
- That is, we cannot say “We are 95% confident that there is no relationship” if we fail to reject H0!