

Hypothesis testing

Brian Healy, PhD

Harvard Catalyst Applied
Biostatistics Certificate Program

Previous classes

- Basics of probability
- Central limit theorem
- Sample mean

Lecture outline

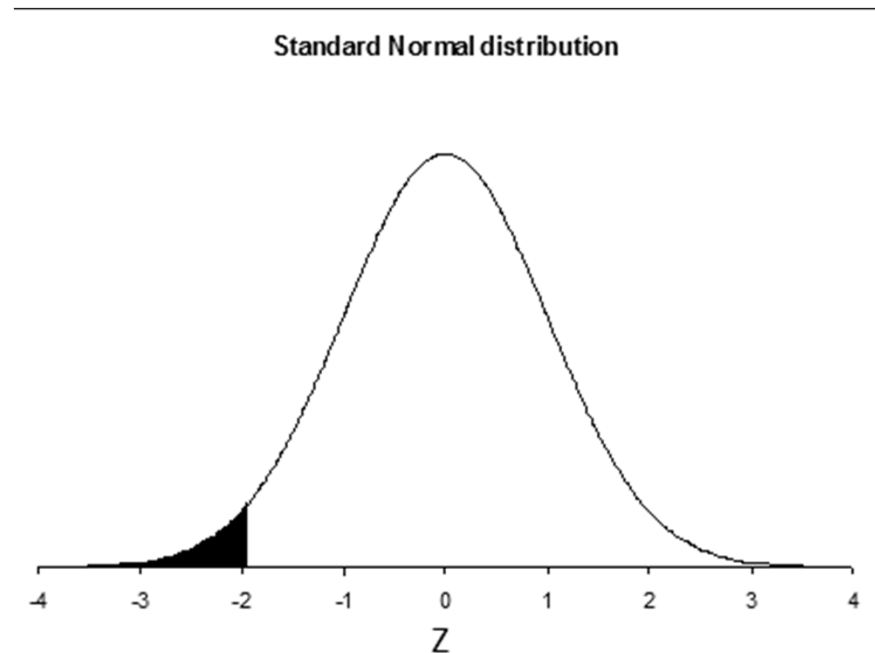
- Rare event
- p-value
- Hypothesis test
- One sample and two sample t-test
- Confidence intervals

Goals of lecture

- At the end of this lecture, you will:
 - Understand the concept of the p-value
 - Understand the steps involved in a hypothesis test
 - Understand the concept of a confidence interval
 - Be able to interpret the output from a hypothesis test in STATA

Review

- In lecture 2, we investigate the probability of specific events under a normal distribution
- $P(Z \leq -1.96) = 0.025$

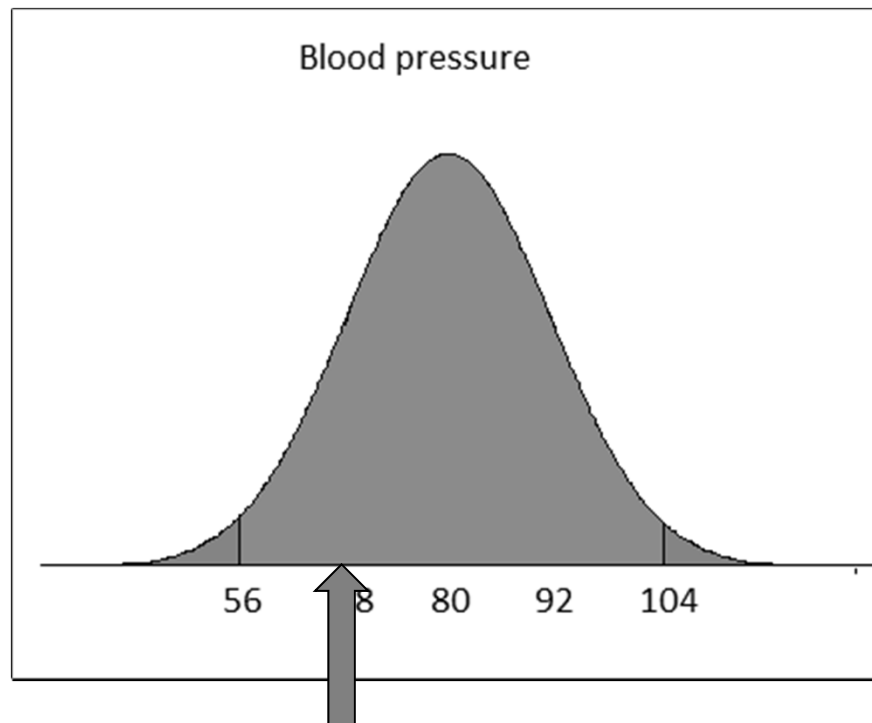


Bounds

- What if we wanted to know the normal range for blood pressure?
 - Let's define normal range to be the middle 95% of the data
 - We know $P(-1.96 \leq Z \leq 1.96) = 0.95$
 - $P\left(-1.96 \leq \frac{X-80}{12} \leq 1.96\right) =$
 - $P(80 - 1.96 * 12 \leq X \leq 80 + 1.96 * 12) =$
 - $P(56.5 \leq X \leq 103.5) = 0.95$

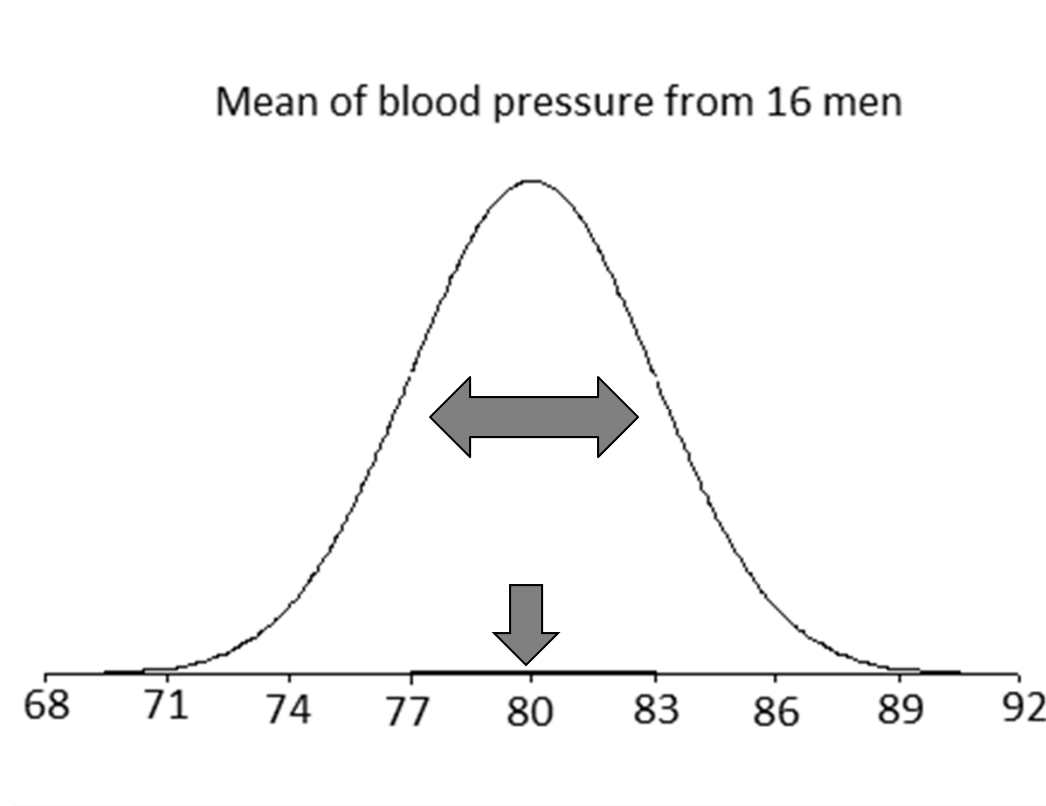
Bounds

- 95% of the data fall between (56.5, 103.5)



95% of the data fall
in the green area

Distribution of sample mean



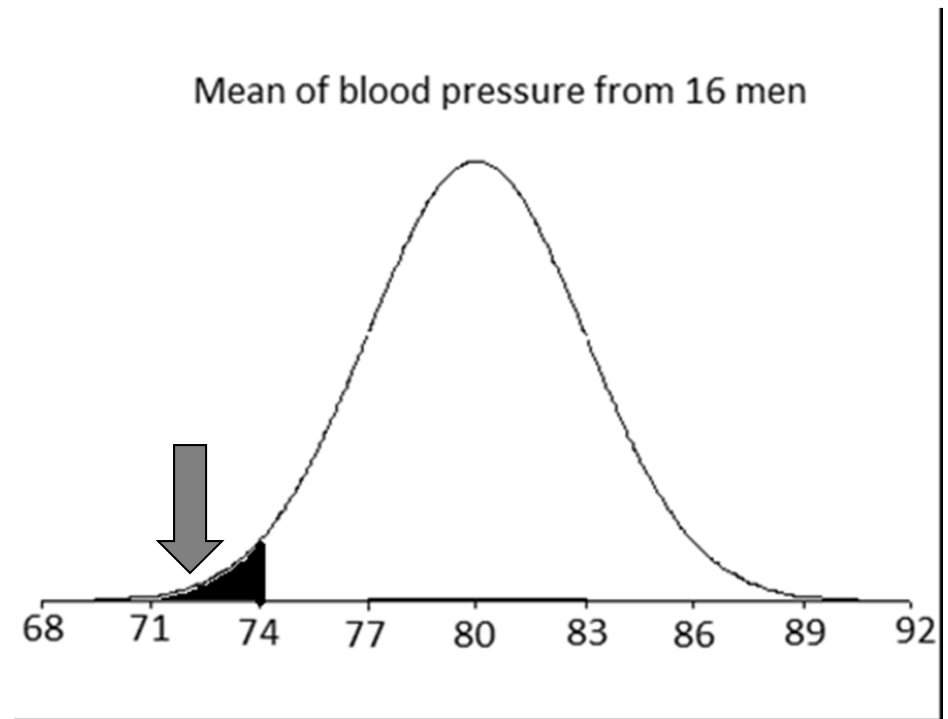
Distribution of sample means in normal

Mean is equal to the population mean=80

Standard deviation is equal to population standard deviation divided by square root of sample size=3

Review

- To calculate the probability of the mean of 16 men being less than 74, we can use the following expression



$$P(\bar{X} \leq 74) = P\left(\frac{\bar{X} - 80}{12/\sqrt{16}} \leq \frac{74 - 80}{12/\sqrt{16}}\right) = P(Z \leq 2) = 0.023$$

Conclusions

- When we know the distribution of a statistic, we can calculate the probability of specific events
- If data follow a normal distribution, we can calculate the probability of being less than a specific value using the standard normal distribution
 - This is called the tail probability

Big picture

- Last week, we discussed how to calculate specific probabilities when we knew parameters of the distribution
- Most of the time we do not know the parameters of a distribution
 - If we knew the mean of a population, we would not need to do a study
- Rather, we would like to learn about a new group
 - Hypothesis test
 - Confidence interval

Example

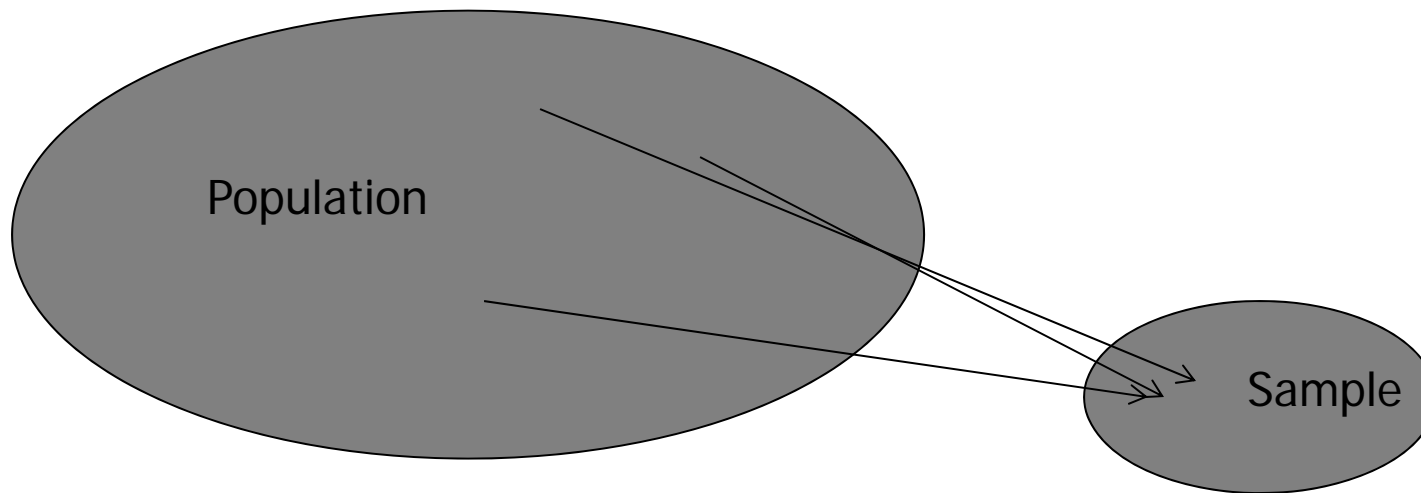
- Multiple sclerosis (MS) is the most common neurologic disorder among young people in the US
- Although MS primarily leads to physical disability, cognitive difficulties have been increasingly recognized
- Is the average cognitive functioning in patients with relapsing-remitting MS different than healthy controls?

SDMT

- One way to measure speed of information processing (a domain of cognitive functioning) is the symbol digit modalities test (SDMT)
- We would like to investigate the mean SDMT score in RRMS patients
 - Option 1: Measure the SDMT in ALL patients and compare it to mean in healthy controls
 - Option 2: Measure the SDMT in a sample of patients and compare it to mean in healthy controls

Sample

- Remember we would like our sample to be
 - Representative of the population
 - Random



How do we test this hypothesis?

- In the normal population, the mean SDMT score is 55
- How could we determine if RRMS patients are different than the normal population?
 - Hypothesis test
- Take a sample
 - Sample mean = 54.5
 - Is this equal to the hypothesized mean?
 - NO!!!
 - SUCCESS!!! We are done!!!

Reasons for differences between groups

- Actual effect-when there is a difference between the two groups
- **Chance**
- **Bias**
- **Confounding**
- Statistical tests are designed to determine if the observed difference between the groups was likely due to chance

Chance experiment

- Experiment: I flip a coin
 - If heads, I win \$1
 - If tails, you win \$1
- What if the following happened?
 - 2 heads in a row
 - 5 heads in a row
 - 15 heads in a row
- When are you suspicious?

Null hypothesis

- In all experiments, we have an initial belief
 - In coin example, you believed that there was a 50/50 chance of heads
- We always set up our null hypothesis so that we can reject the null hypothesis.
- For our study, the null hypothesis is that the mean SDMT in the RRMS patients is equal to 55.
 - H_0 : Mean SDMT=55
 - H_0 : $\mu_0=55$

Alternative hypothesis

- Notation: H_A or H_1
- Has two characteristics
 - Must cover all values not included in the null
 - Must contain the value that we think is going to happen
- H_A : Mean $SDMT_{RRMS}$ is not equal to 55
- H_A : The probability of heads is not 0.5

Hypothesis test

- Definition: A statistical test of a null hypothesis
- Completed under the assumption that the null is true (conditional probability)
- Always want to disprove the null hypothesis
 - Ex. 1- H_0 : Mean $SDMT_{RRMS}=55$
 - H_A : Mean $SDMT_{RRMS} \neq 55$
 - Ex. 2- H_0 : Probability of heads=0.5
 - H_A : Probability of heads $\neq 0.5$
- The most important step is properly defining the null and alternative hypotheses

P-value

- **Definition:** the probability of the observed result or something more extreme under the null hypothesis
- If the probability of the event is sufficiently small, we say that the difference is likely not due simply to chance and we have an actual effect.
- If p-value is small enough, we call the effect **statistically significant**

What is small enough?

- Scientific community has determined that 0.05 is small enough
 - 0.05 corresponds to a 1 out of 20 chance
- We will reject our null hypothesis if the event is unlikely under the null
- **Type I error** is the probability that we reject the null even though the null is true
 - More later

Aside-Bias

- Is there something in my design that led to my result?
- Do I have a random and representative sample of the group that I meant to study?
 - Extremely important to consider bias when doing a one-sample hypothesis test

Steps for hypothesis testing

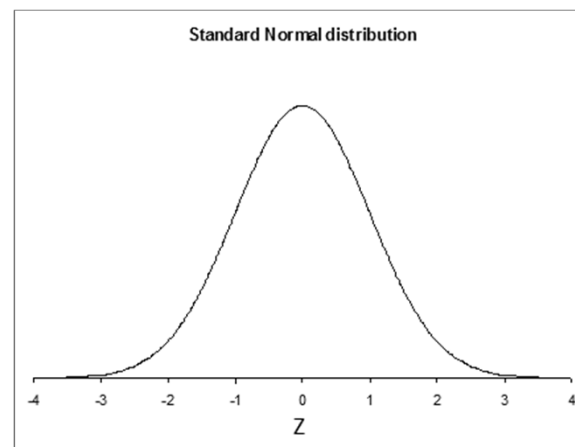
- 1) State null hypothesis
- 2) State type of data
- 3) Determine appropriate statistical test
- 4) State summary statistics if possible
- 5) Calculate p-value (stat package)
- 6) Decide whether to reject or not reject the null hypothesis
- 7) Write conclusion

Example

- 1) $H_0: \text{mean}_{\text{RRMS}} = 55$ ($\mu_0=55$)
- 2) Outcome: SDMT score-continuous
 - How different is our observed data from the hypothesized value?
 - What test can we use to compare the mean of a continuous outcome to a known value?

Distribution of mean under the null

- Initially, let's assume that we know the standard deviation is 12 and that the data come from a normal distribution
- Under the null hypothesis, the mean=55
- Given this information, we know the distribution of the sample mean under the null hypothesis

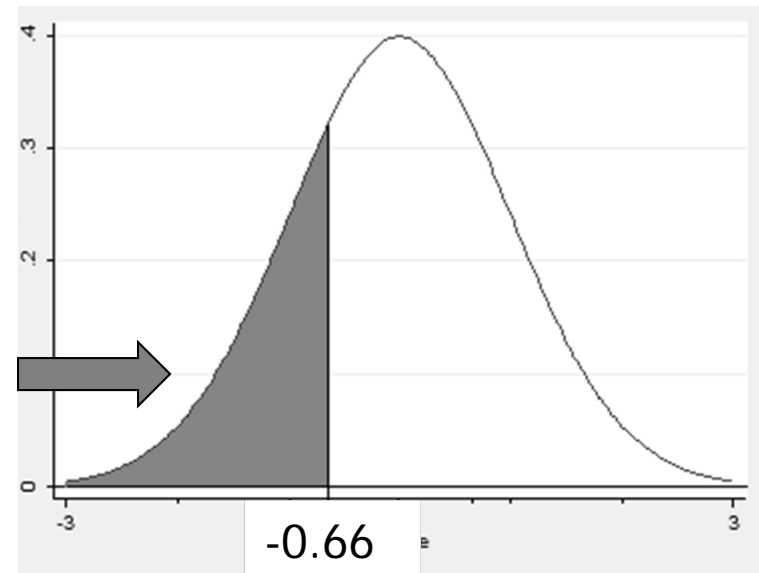


$$Z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

Is the observed mean different from the hypothesized mean relative to the variability in the data?

Picture

- Our observed z-statistic is $z = \frac{54.5 - 55}{12.1 / \sqrt{252}} = -0.66$
- To calculate the p-value, we calculate the probability of the observed z-statistic or something more extreme
- Area in lower tail



Picture

- Using the concept we learned in Lecture 2, we calculate the area in the lower tail, which in math terms is $P(Z \leq -0.66)$
- Using STATA, we calculate $P(Z \leq -0.66) = 0.25$

- Using the concept we learned in Lecture 2, we calculate the area in the lower tail, which in math terms is $P(Z \leq -0.66)$
- Using STATA, we calculate $P(Z \leq -0.66) = 0.25$

-0.66

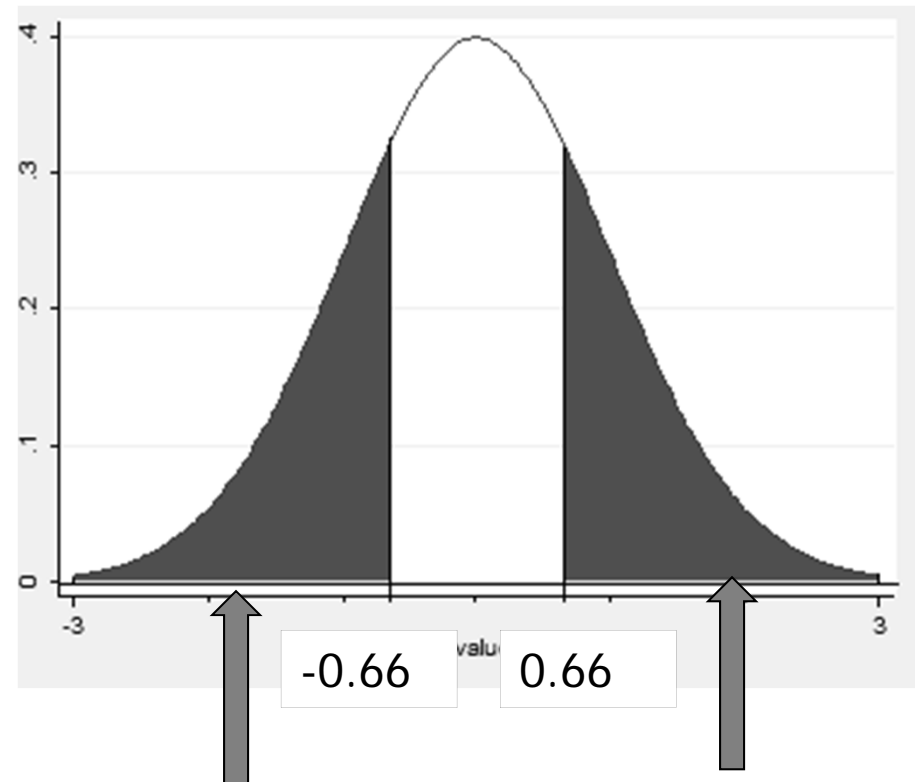
The area in this tail is 0.25

Amazing!!!

- Remember that the p-value is the probability of the observed result or something more extreme under the null
- On the previous slide, we calculated the probability of our observed z-statistic or something more extreme under the null
- This is our one-sided p-value!!
- Since our p-value is 0.25, what does that tell us?

Two-sided p-value

- In medical statistics, the vast majority of the time we are interested in two sided p-values
- This means that our p-value is the probability of the observed result or something more extreme in both directions
- Two-sided p-value = $2 \times$ one-sided p-value



The area in this tail is 0.25

The area in this tail is 0.25

Results

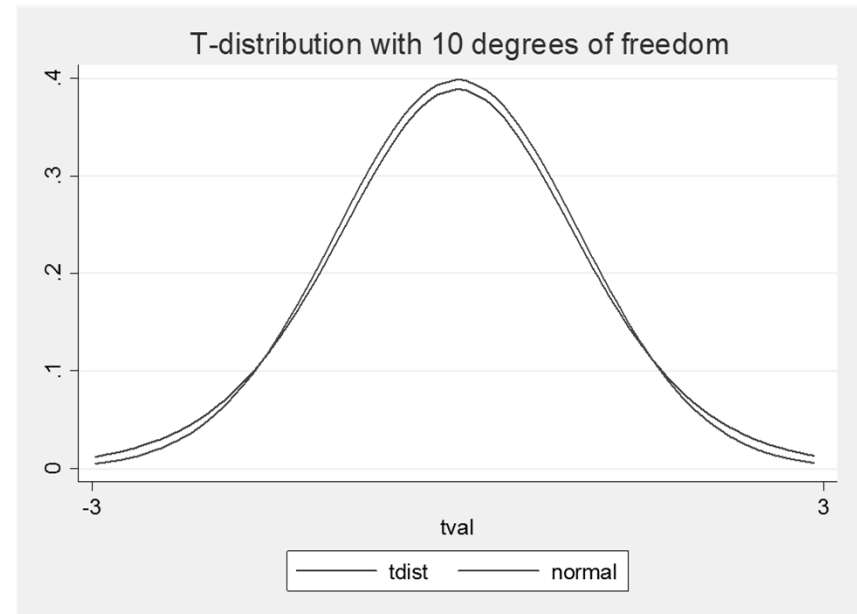
- 1) $H_0: \text{mean}_{\text{RRMS}} = 55$
- 2) Outcome: SDMT score-continuous
- 3) Test: One sample z-test
- 4) Estimated mean ($\hat{\text{mean}}_{\text{RRMS}}$) = 54.5
- 5) Calculate p-value: 0.51
- 6) Fail to reject the null hypothesis because p-value is greater than 0.05
- 7) Conclusion: The mean SDMT score in RRMS patients is not significantly different than 55.

One sample t-test

- All of the previous results assumed that we knew the standard deviation
- If we do not know the standard deviation, we can use the sample standard deviation by switching to a t-test
 - Much more commonly used
 - **Assumption: Underlying population distribution is normal and independent observations**
- The test assesses whether the observed mean is different from what we would expect under the null

t-distribution

- The t-distribution is very similar to the normal distribution, but it has slightly heavier tails
- Confidence intervals using a t-distribution are wider than normal distribution confidence intervals



Notes

- The t-distribution and normal become very similar as the sample size increases
- Therefore, hypothesis tests in large samples can be based on the normal or the t-distribution without much change even when the SD is estimated

Degrees of freedom	t-value for 95% CI
1	12.7
10	2.23
100	1.98
1000	1.96

Test statistic

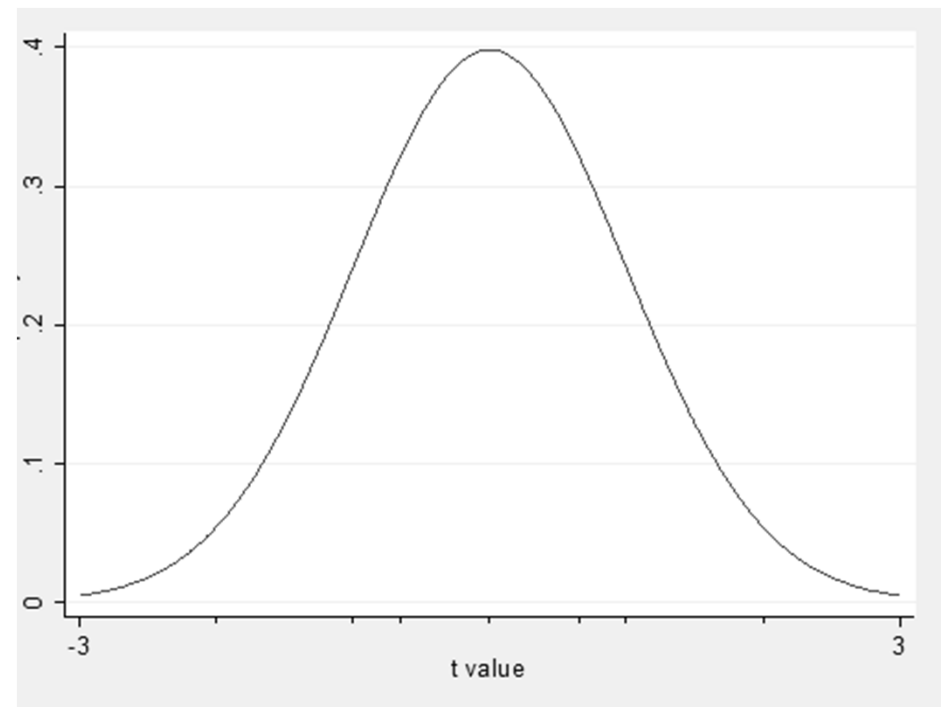
- Since we are now estimating the standard deviation rather than assuming it is known, the statistic that we calculate changes slightly
- Basic form of test statistic:

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

Is the observed mean different from the hypothesized mean relative to the variability in the data?

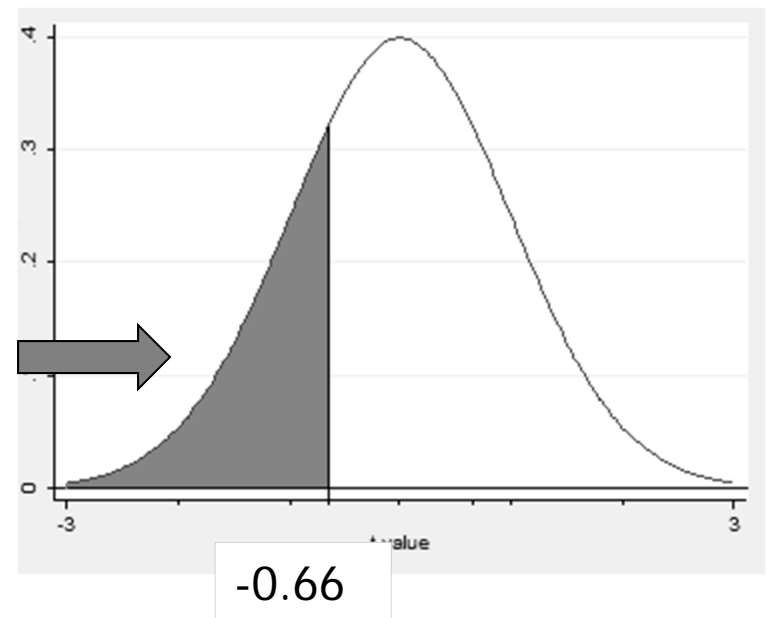
Picture

- If we repeatedly sampled people from a normal distribution with mean=55, the calculated t-statistics from these sample would have this distribution (i.e. this is the distribution under the null hypothesis)



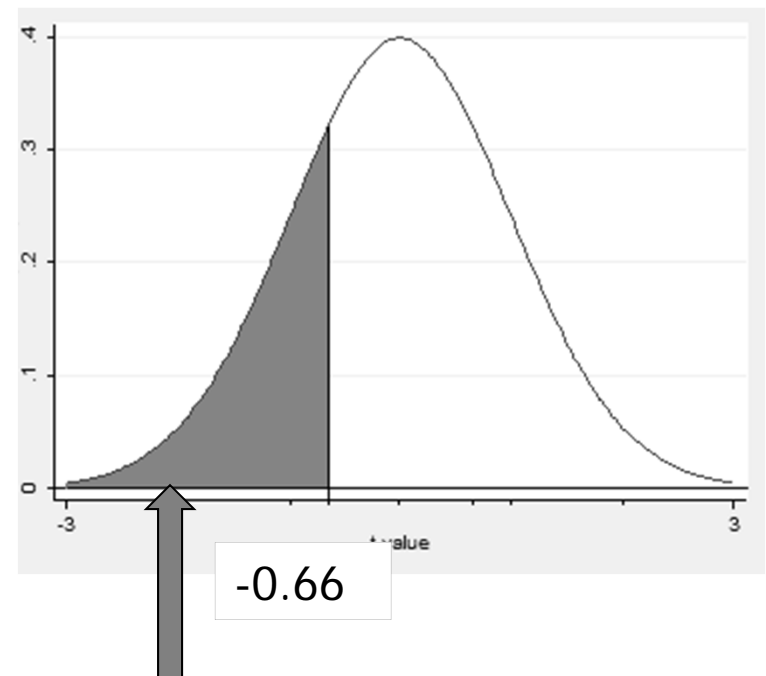
Picture

- Our observed t-statistic is $t = \frac{54.5 - 55}{12.1 / \sqrt{252}} = -0.66$
- To calculate the p-value, we calculate the probability of the observed t-statistic or something more extreme
- Area in lower tail



Picture

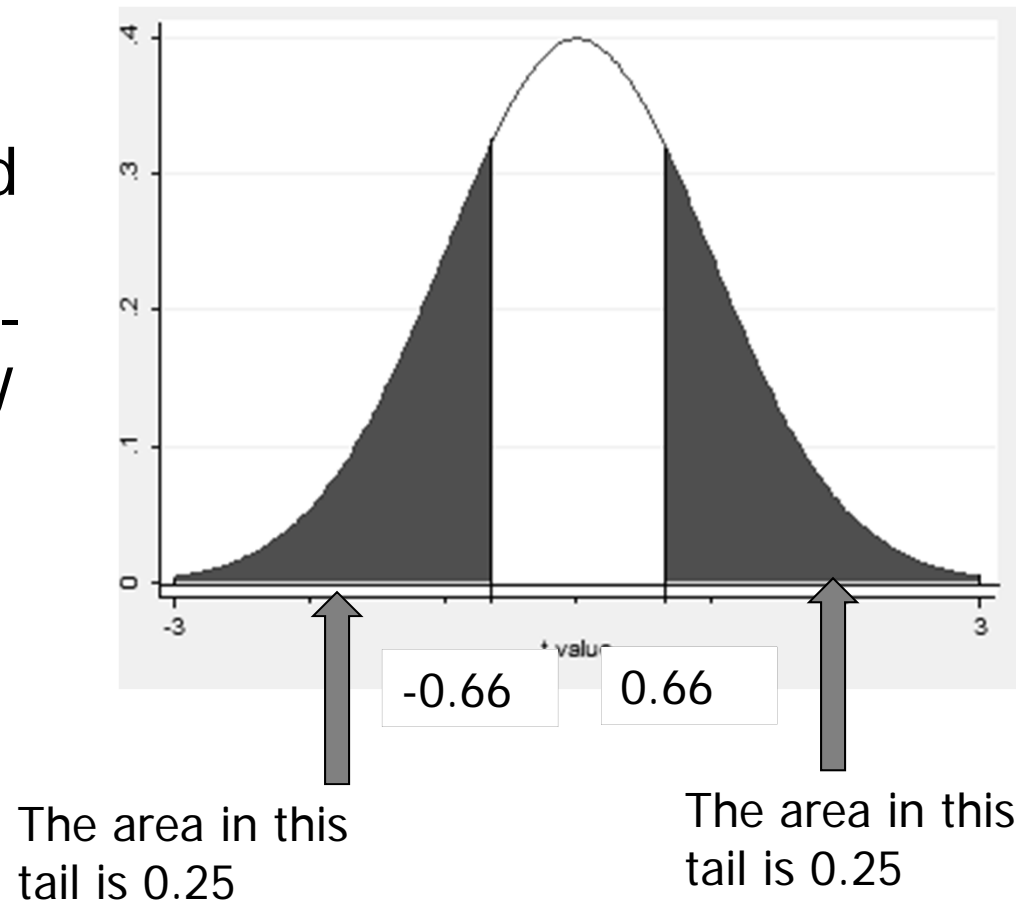
- Using the concept we learned in Lecture 2, we calculate the area in the lower tail, which in math terms is $P(T \leq -0.66)$
- Using STATA, we calculate $P(T \leq -0.66) = 0.25$



The area in this tail is 0.25

Two-sided p-value

- In medical statistics, the vast majority of the time we are interested in two sided p-values
- This means that our p-value is the probability of the observed result or something more extreme in both directions
- Two-sided p-value = $2 \times$ one-sided p-value



Results

- 1) $H_0: \text{mean}_{\text{RRMS}} = 55$
- 2) Outcome: SDMT score-continuous
- 3) Test: One sample t-test
- 4) Estimated mean ($\hat{\text{mean}}_{\text{RRMS}}$) = 54.5
- 5) Calculate p-value: 0.51
- 6) Fail to reject the null hypothesis because p-value is greater than 0.05
- 7) Conclusion: The mean SDMT score in RRMS patients is not significantly different than 55.

STATA output

STATA command

`. ttest sdmtfin == 55`

One-sample t test

Variable	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
sdmtfin	252	54.49512	.7598783	12.06269	52.99857	55.99167

t-statistic

mean = mean(sdmtfin)
Ho: mean = 55

t = -0.6644
degrees of freedom = 251

Ha: mean < 55
Pr(T < t) = 0.2535

Ha: mean != 55
Pr(|T| > |t|) = 0.5070

Ha: mean > 55
Pr(T > t) = 0.7465

One-sided p-value

Two-sided p-value

Comparison of one-sided and two-sided tests

- Two-sided p-value is twice one-sided p-value for a t-test
- Two-sided test is more conservative because the rejection region is split between the high and low side. For the one-sided test, the rejection region is only on the side of interest
- Two-sided test most common in literature even though people sometimes know the direction of effect they are interested in detecting.

What if $p > 0.05$?

- In this case, the difference between the mean in the RRMS and the hypothesized value is not statistically significant (at the 0.05 level).
- “If observed mean is not significantly different from the null, then by definition are they not identical?”
 - No
 - The RRMS patients are not significantly different, but we cannot say that they are the same
 - We fail to reject the null hypothesis; we do not accept that the null is true

Summary

- One-sample hypothesis test allows us to compare our sample to a hypothesized value
- Based on our hypothesis test or the confidence interval, we conclude that there is no statistical significant difference between RRMS patients and healthy controls

Conclusions

- Most important step of hypothesis test is deciding the null hypothesis
- To complete a hypothesis test, we determine the probability of the observed result or something more extreme under the null
- If observed result or something more extreme is sufficiently rare, we reject null

Estimation of effect

- Although hypothesis tests are a critical part of biostatistics, the complimentary part is the estimate of the effect
- Beyond simply if there is a significant difference between healthy controls and RRMS, we would like to know an estimate of the performance in RRMS
- The sample mean provides a point estimate of the population mean
 - Sample mean=54.5

Something more

- Although we have an estimate of the population mean, can we incorporate the accuracy of our estimate?
- Can we use both the point estimate and the information about the variability in the sample mean to give more information?

Confidence interval

- Definition: a set of values that we believe are plausible estimates of population mean based on the sample we have drawn
- The confidence interval is an interval around our sample mean that allows us to have a certain amount of confidence that the population mean is covered by the interval.

Construction of a confidence interval

- To construct a confidence interval we need to go back to probability.
- We are going to start with a standard normal. Remember that 1.96 leaves 0.025 in the upper tail.

$$P(-1.96 < Z < 1.96) = 0.95$$

$$P(-1.96 < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < 1.96) = 0.95$$

$$P(\bar{X} - 1.96 \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + 1.96 \frac{\sigma}{\sqrt{n}}) = 0.95$$

- The resulting interval we get is,

$$\left(\bar{X} - 1.96 * \frac{\sigma}{\sqrt{n}}, \bar{X} + 1.96 * \frac{\sigma}{\sqrt{n}} \right)$$

which is interpreted as that we are 95% confident that this interval will cover μ

- We must be careful about the interpretation:
 - This means “if we selected 100 random samples from the population and calculated 100 confidence intervals for μ , approximately 95 of the intervals would cover μ and 5 would not” (from Pagano and Gauvreau, *Principles of Biostatistics*)
- A more general confidence interval is

$$\left(\bar{X} - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \right)$$

Confidence interval with the t-distribution

- Usually, we do not know the population standard deviation
- If underlying data is normal, we can use a t-distribution even with a small sample
- The entire procedure for finding the confidence interval is the same as for the normal confidence interval, but the cut-offs are from the t-distribution.

$$\left(\bar{X} - t_{n-1, 1-\alpha/2} \frac{s}{\sqrt{n}}, \bar{X} + t_{n-1, 1-\alpha/2} \frac{s}{\sqrt{n}} \right)$$

Concept-technical

- It is important to remember that the population mean is not a random variable. Rather, the interval is the random variable that is subject to variability
 - This is why we do not say that there is a 95% chance that the population mean is in the interval
 - Rather we say we are 95% confident that the interval covers the true mean

Illustration

- Let's look at this through a simulation
- http://www.ruf.rice.edu/~lane/stat_sim/conf_interval/
- Note that we do not always cover the population mean exactly 95 times out of 100, but on average we will.
- What can you say about the 95% and 99% confidence intervals?

Changing the width of the confidence interval

- The width of the confidence interval is based on 3 factors
 - confidence level (z or t)- how confident do we want to be that the interval covers μ
 - variance (σ)- how different are the people
 - sample size (n)- how many samples did we use to estimate the population mean
- Which of these is under the control of the investigator?

Example

- Here we use a t-distribution
- We would like to provide a 95% confidence interval for the mean SDMT score in RRMS patients

$$\left(54.5 - 1.97 \frac{12.1}{\sqrt{252}}, 54.5 + 1.97 \frac{12.1}{\sqrt{252}} \right) = (53.0, 56.0)$$

- For a 99% interval, notice that increasing the confidence level requires increasing the width of the interval as well.

$$\left(54.5 - 2.60 \frac{12.1}{\sqrt{252}}, 54.5 + 2.60 \frac{12.1}{\sqrt{252}} \right) = (52.5, 56.5)$$

Example 2

- Notice that if we decrease the sample size from 252 to 50 and everything else remained the same, the confidence interval would become wider

$$\left(54.5 - 1.97 \frac{12.1}{\sqrt{50}}, 54.5 + 1.97 \frac{12.1}{\sqrt{50}} \right) = (51.1, 57.9)$$

- Finally, notice if the variance decreases, the confidence interval narrows

$$\left(54.5 - 1.97 \frac{7.5}{\sqrt{252}}, 54.5 + 1.97 \frac{7.5}{\sqrt{252}} \right) = (53.6, 55.4)$$

STATA output

STATA command

`. ttest sdmtfin == 55`

One-sample t test

Variable	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
sdmtfin	252	54.49512	.7598783	12.06269	52.99857	55.99167

t-statistic

mean = mean(sdmtfin)
Ho: mean = 55

t = -0.6644
degrees of freedom = 251

Ha: mean < 55
Pr(T < t) = 0.2535

Ha: mean != 55
Pr(|T| > |t|) = 0.5070

Ha: mean > 55
Pr(T > t) = 0.7465

One-sided p-value

Two-sided p-value

Comparison to hypothesis test

- We calculated a 95% confidence interval for the mean score based on our sample (53.0, 56.0)
- Since the null value of 55 is included in the confidence interval, this is one of the plausible values
 - Since the null is a plausible value, we fail to reject the null hypothesis
- The conclusion we reach using the CI and the hypothesis test match

Conclusions

- We are 95% confident that the true mean level of SDMT lies between (53.0,56.0)
- Remember that our confidence interval is subject to the sampling variability so that sometimes just by chance our confidence interval will not cover the true population mean.

One-sided confidence interval

- One-sided CIs are very common in everyday life (catching a bus or train), but far less common in medical applications
- These are either a lower (upper) bound because instead of being 95% confident that the mean is in an interval, we now say that we are 95% confident that the mean is above (below) a given value. To have 0.05 in the lower (upper) tail, the cut-off from the standard normal distribution is -1.645 (1.645).

One-sided continued

- The 95% one-sided confidence interval (lower bound) is $\left(\bar{X} - 1.645 \frac{\sigma}{\sqrt{n}}, \infty\right)$ or $\left(\bar{X} - t_{n-1, 0.95} \frac{s}{\sqrt{n}}, \infty\right)$
- The 95% one-sided confidence interval (upper bound) is $\left(-\infty, \bar{X} + 1.645 \frac{\sigma}{\sqrt{n}}\right)$ or $\left(-\infty, \bar{X} + t_{n-1, 0.95} \frac{s}{\sqrt{n}}\right)$
- The interpretation of these are “we are 95% confident that μ is at most (at least) the given upper bound (lower bound).”

Summary

- A major goal in medical research is to estimate the mean of a population based on a sample
 - Sample mean provides the best point estimate
 - Variability in the data (either known or estimated from the sample) allows you to provide a range of plausible values through the confidence interval
- Confidence interval and hypothesis test provide complimentary information

Two sample t-test

Comparison of two groups

- Although we sometimes compare one group to a known population value, it is much more common to compare two groups
- Example:
 - Clinical trials (Placebo vs. Treated)
 - Observational studies (Exposed vs. Unexposed)

Example

- Two main forms of MS
 - Relapsing
 - Progressive
- MS is known to be related to the immune system
- Question: Is immune system involvement different in two forms of MS?
 - Marker: CD-26

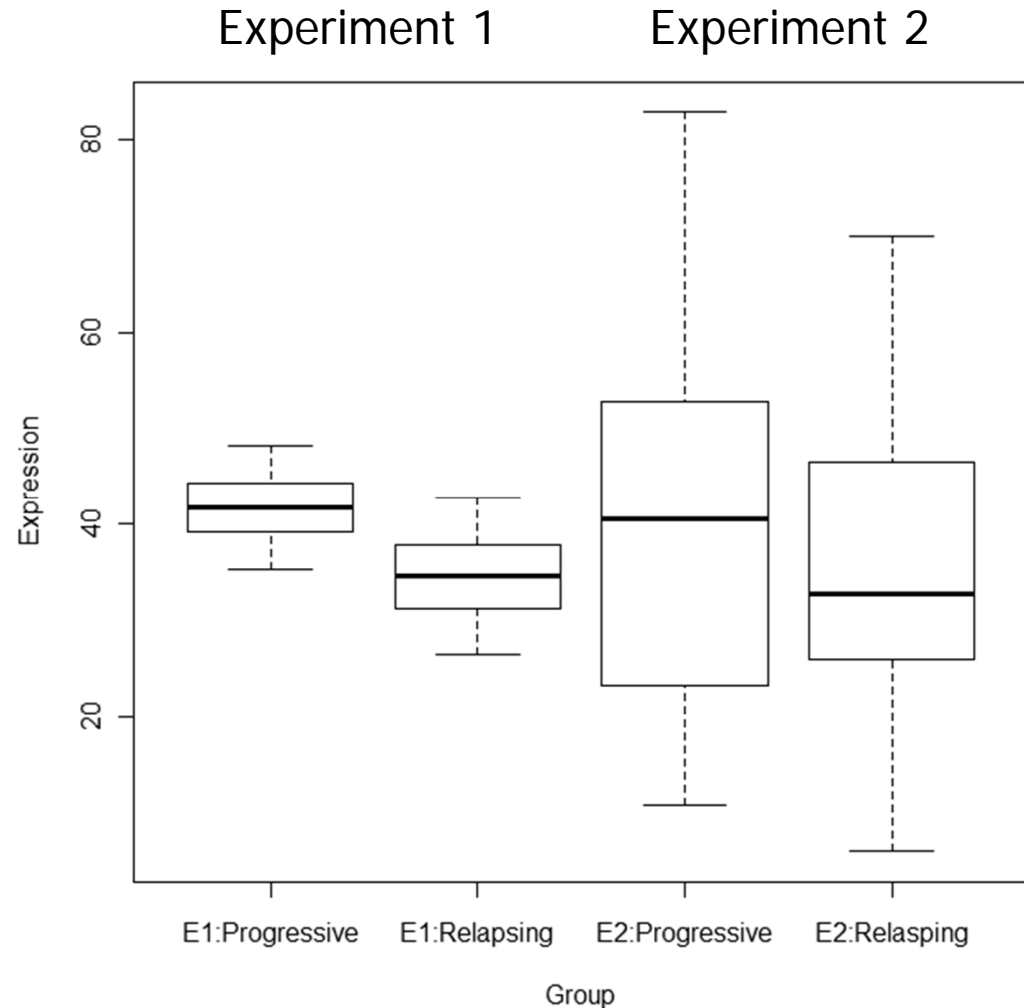
Example

- Question: Is the expression of CD-26 different in relapsing MS patients compared to progressive MS patients?
- What is the outcome?
 - We measure CD-26 using flow cytometry
 - Continuous variable
- What is the explanatory variable?
 - Group membership (relapsing vs. progressive)
 - Dichotomous variable
- How would you answer this question?
 - Collect a sample from each group

Results

- Mean values in our sample:
 - Estimated mean (mêan)
 - Relapsing patients=37.4
 - Progressive patients=44.3
 - Does the mean in the sample exactly equal the mean in the population?
- The progressive patients had greater production, but are we certain that there is a difference between these?
 - **Statistically significant**
 - Clinically meaningful
- What is the variability in the data?

- Means in two groups are the same in both experiments
- Is there a difference in Experiment 1?
- In Experiment 2?
- Hypothesis test



Example

1) $H_0: \text{mean}_{\text{relapsing}} = \text{mean}_{\text{progressive}}$

$$H_0: \text{mean}_{\text{relapsing}} - \text{mean}_{\text{progressive}} = 0$$

2) Explanatory: group membership-dichotomous

Outcome: cytokine production-continuous

- What test can we use to compare a continuous outcome with a dichotomous explanatory variable?

Two sample t-test

- A two sample t-test is a test for differences in means in two samples.
- **Assumption: Underlying population distribution is normal and independent observations**
- To calculate p-value, we need to understand if the difference in the means is large relative to the variability in data
- Can get p-value from statistical package

Difference in the sample means

- We are going to use the difference of the means for our hypothesis test, but we need to estimate the variance of this difference to determine if the difference is significant
- Basic form of test statistic:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_{\bar{x}_1 - \bar{x}_2}}$$

Is the observed mean difference similar to the hypothesized mean difference relative to the variability in the data?

Equal variance

- Sometimes we will be willing to assume that the variance in the two groups is equal:

$$\sigma_1^2 = \sigma_2^2 = \sigma^2$$

- We have an estimate of σ^2 from each of the samples in the sample variances
- Since we have two estimates of one quantity, we pool the two estimates

Equal variance continued

- The estimate of σ^2 is given by:

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

- The t-statistic based on the pooled variance:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

- The t-statistic has a t-distribution with $n_1 + n_2 - 2$ degrees of freedom

Unequal variance

- Often, we are unwilling to assume that the variances are equal
- We now write the test statistic as:

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

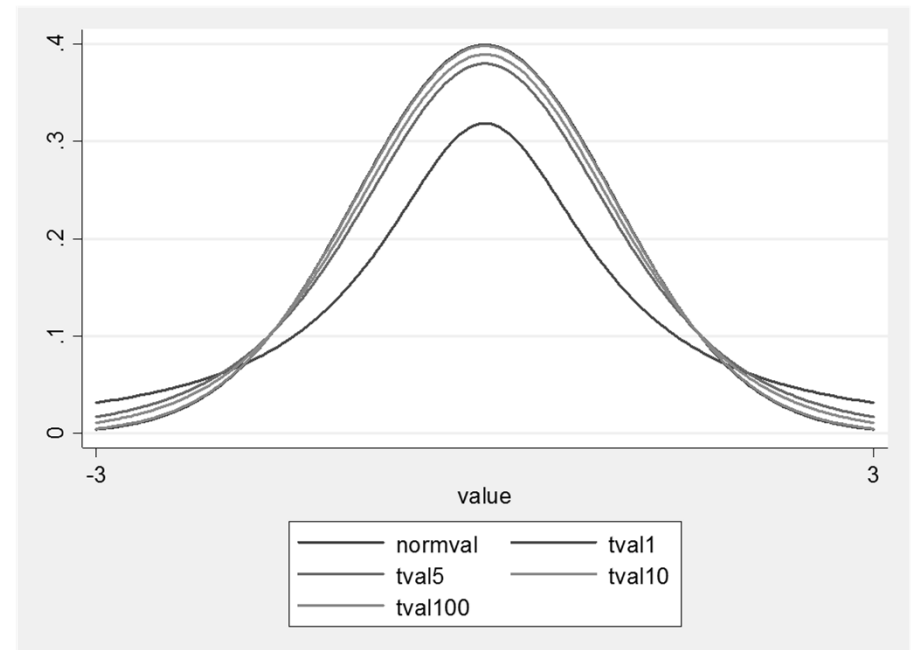
- The distribution of this statistic is difficult to derive and we approximate the distribution using a t-distribution with ν degrees of freedom

$$\nu = \frac{\left[\left(\frac{s_1^2}{n_1} \right) + \left(\frac{s_2^2}{n_2} \right) \right]^2}{\left[\frac{\left(\frac{s_1^2}{n_1} \right)^2}{(n_1 - 1)} + \frac{\left(\frac{s_2^2}{n_2} \right)^2}{(n_2 - 1)} \right]}$$

- This is called the Satterthwaite approximation
 - When you complete a two-sample t-test in STATA and the variances are not assumed equal, this approximation is used

Degrees of freedom

- Degrees of freedom for a t-test are related to the sample size
- As degrees of freedom increases for t-test, the shape becomes more like a normal



Hypothesis test

- 1) $H_0: \text{mean}_{\text{relapsing}} = \text{mean}_{\text{progressive}}$
 $H_0: \text{mean}_{\text{relapsing}} - \text{mean}_{\text{progressive}} = 0$
- 2) Explanatory: dichotomous, outcome: continuous
- 3) Two sample t-test
- 4) Estimated mean ($\hat{\text{mean}}_{\text{relapsing}}$) = 37.4,
Estimated mean ($\hat{\text{mean}}_{\text{progressive}}$) = 44.3
- 4) Calculate p-value: $p = 0.28$
- 6) Fail to reject the null hypothesis because p-value is greater than 0.05
- 7) Conclusion: The difference between the groups is not statistically significant.

Summary statistics

Group specific confidence intervals

Estimated mean group difference

. ttest expression, by(group) unequal
Two-sample t test with unequal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
Progress	17	44.25647	5.61928	23.16888	32.34413	56.16881
Relapsin	34	37.38735	2.47182	14.41307	32.3584	42.41631
combined	51	39.67706	2.503239	17.8767	34.64916	44.70496
diff		6.869118	6.138909		-5.84953	19.58777

diff = mean(Progress) - mean(Relapsin) t = 1.1189
Ho: diff = 0 Satterthwaite's degrees of freedom = 22.3846

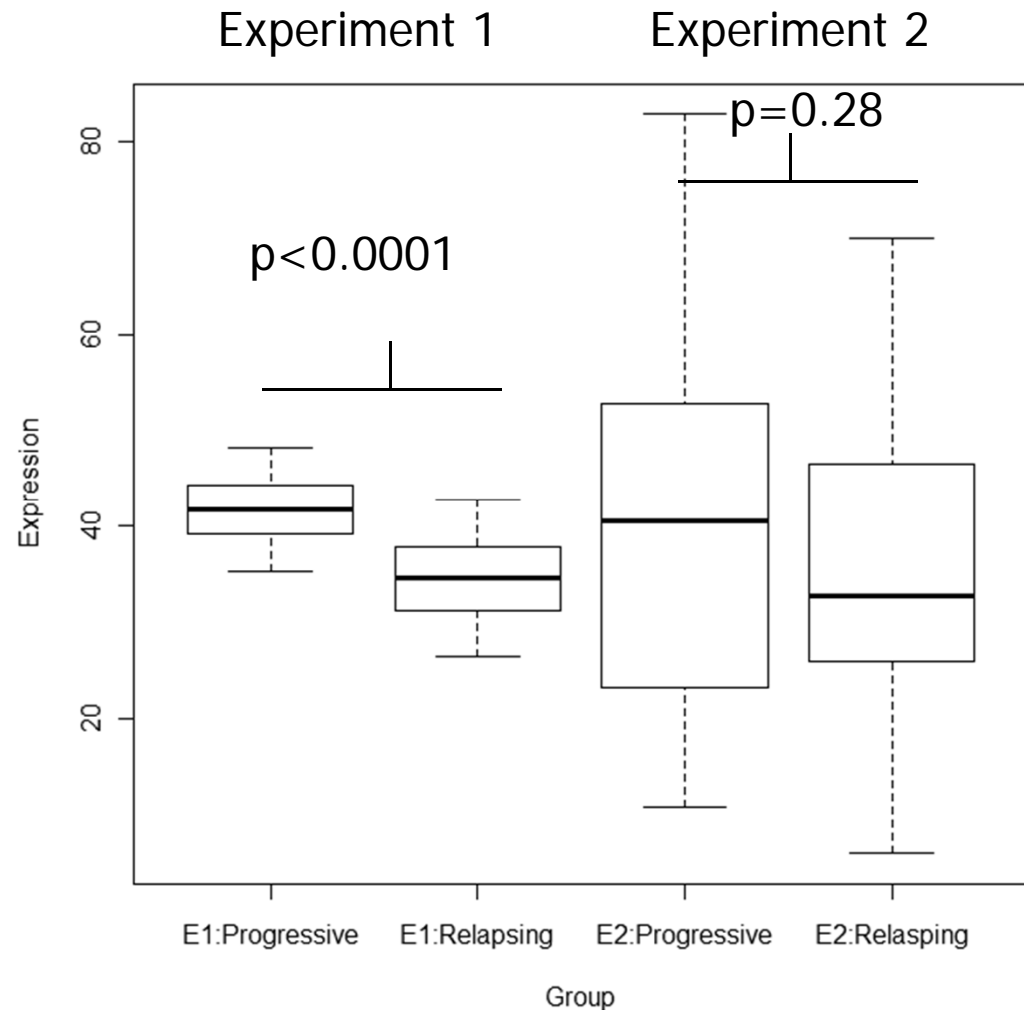
Ha: diff < 0 Pr(T < t) = 0.8625	Ha: diff != 0 Pr(T > t) = 0.2750	Ha: diff > 0 Pr(T > t) = 0.1375
------------------------------------	-----------------------------------------	------------------------------------

95% CI for mean group difference

Two-sided p-value

One-sided p-value

- Significant difference in experiment 1
- Added variance in experiment 2 led to non-significant result
- What does this mean?



Confidence interval

- Another way to investigate our hypothesis is to use the confidence interval
- For this problem, the difference in the mean between the groups is of interest
 - Estimated difference between the groups = 6.9
 - 95% CI: (-5.8, 19.6)
- What number is included in the confidence interval?
- What does this mean?

Summary

- When we would like to compare the mean of a continuous outcome in two groups, we use a two-sample t-test
- To determine if the groups are significantly different, we calculate a p-value or use the 95% confidence interval

Assumptions

Equal vs. unequal variance

- Generally, I use the unequal variance t-test as my default (this is the default in some stat packages)
- Best to choose which type of test you will use prior to looking at the data and the results
- Can also test for equal variance
 - $H_0: \sigma_1^2 = \sigma_2^2$
 - $H_A: \sigma_1^2 \neq \sigma_2^2$

Test for equal variance

- In this case, rejecting the null hypothesis indicates that the variances are not equal
- Failing to reject the null indicates that the assumption of equal variance seems reasonable
- Is lack of equal variance interesting?
- STATA provides a test for equal variance, but this test is sensitive to normality
- Limited power

STATA output

■ Here is the STATA output

```
. sdtest expression, by(group)
```

Variance ratio test

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
Progress Relapsin	17	44.25647	5.61928	23.16888	32.34413	56.16881
	34	37.38735	2.47182	14.41307	32.3584	42.41631
combined	51	39.67706	2.503239	17.8767	34.64916	44.70496

ratio = sd(Progress) / sd(Relapsin) f = 2.5840
Ho: ratio = 1 degrees of freedom = 16, 33

Ha: ratio < 1
Pr(F < f) = 0.9897

Ha: ratio != 1
2*Pr(F > f) = 0.0207

Ha: ratio > 1
Pr(F > f) = 0.0103

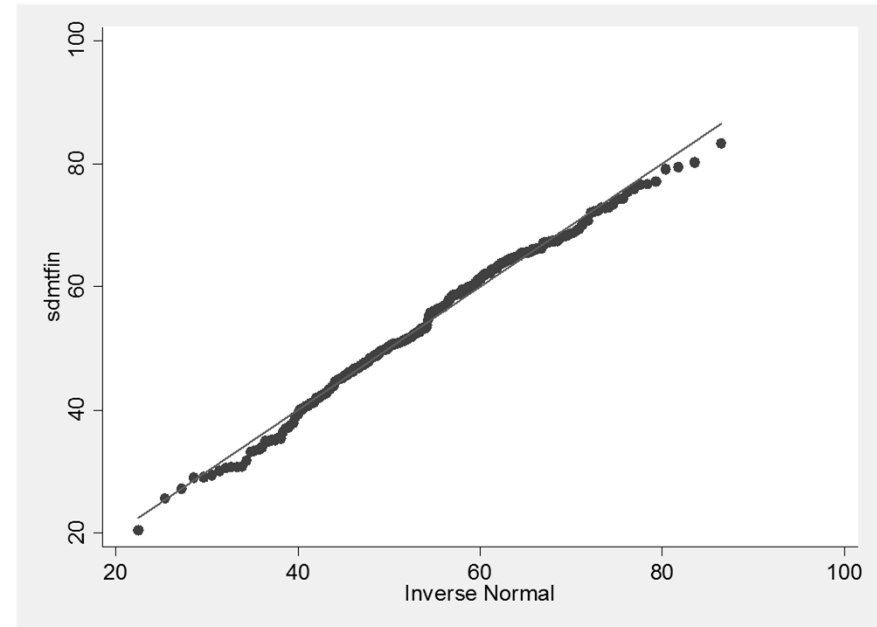
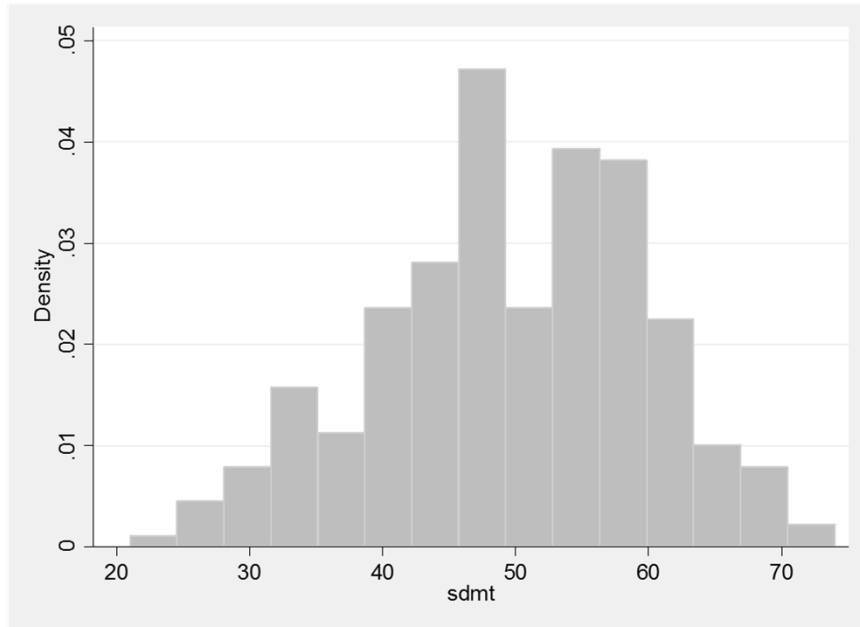


Two-sided p-value

Normality

- Both t-test also makes an assumption of normality of underlying data
- The t-test is very **robust** to departures from this assumption with a reasonable sample size ($n \geq 50$)
- Ways to assess normality
 - Inspection of plot of data
 - Formal statistical test
 - My preference are graphical approaches

Graphs



- These are graphs for SDMT and relevant for one-sample t-test
- To inspect the data graphically for two sample t-test, you need to look at each group separately

Test for normality

■ Shapiro-Wilk test

– SDMT example

```
. swilk sdmtfin
```

Shapiro-wilk W test for normal data					
Variable	Obs	W	V	z	Prob>z
sdmtfin	252	0.99036	1.761	1.317	0.09393

← p-value

■ Skewness/kurtosis test

– SDMT example

```
. sktest sdmtfin
```

Skewness/Kurtosis tests for Normality						
Variable	Obs	Pr(Skewness)	Pr(Kurtosis)	adj chi2(2)	joint Prob>chi2	
sdmtfin	252	0.1250	0.1903	4.10	0.1289	

← p-value

What if normality does not hold?

- Transform the data
 - Often a log-transformation can transform a skewed distribution
- If you cannot assume normality, you can change to a nonparametric test
 - Two-sample t-test: Wilcoxon rank sum
 - Permutation test
 - Lecture 5

Conclusions

- Both forms of t-test assume that the data come from an underlying normal distribution
- The t-test is quite robust to departures from normality, especially with moderate/large sample size
- With small sample size, methods that do not require a normality assumption are available

Types of analysis-independent samples

Outcome	Explanatory	Analysis
Continuous	Dichotomous	t-test
Continuous	Categorical	ANOVA, linear regression
Continuous	Continuous	Correlation, linear regression
Ordinal	Dichotomous	Wilcoxon test
Dichotomous	Dichotomous	Chi-square test, logistic regression
Dichotomous	Continuous	Logistic regression
Time to event	Dichotomous	Log-rank test

What you learned

- In this lecture, you learned:
 - Concept of the p-value
 - Steps involved in a hypothesis test
 - How to interpret the STATA output from a t-test