# Biostat 200
# Lecture 5

# Today

- Where are we
- Review of finding percentiles and cutoff values, CLT and 95% confidence intervals
- Hypothesis testing in general
- Hypothesis testing of one mean
- Hypothesis testing of one proportion
- Types of error

# Where are we?

- Week 1: Variables, tables, graphs
- Week 2: Probability:

  Independence → multiplicative rule

  Mutual exclusivity → additive rule

  Conditional probability
- Week 3: Binomial and normal probability distributions, how to get probabilities if you assume these distributions
- Week 4: CLT for the distribution of sample means and proportions and 95% CIs for means and proportions
- Week 5: Hypothesis testing in general and for comparing one mean or one proportion to a hypothesized value
- Weeks 6-8: More hypothesis tests
- Weeks 9-11: Correlation and regression (use hypothesis tests there too)

# Review: Normal distribution

- If $X \sim N(\mu,\sigma)$ then $Z=(X- \mu)/\sigma$
  - To find the proportion of values in the population that exceed some threshold, i.e. $P(X>x)$
    - Transform x to $z=(x - \mu)/\sigma$ and look up $P(Z>z)$ using `display 1-normal(z)`
  - To find the percentiles of the distribution
    - E.g. to find the 97.5 percentile, the value for which 97.5% of the values are smaller, find z by using

      `display invnormal(.975)`

      and solve $z=(x - \mu)/\sigma$ for x

# Review: Normal distribution

- CLT: If X is a random variable with mean μ and standard deviation σ, then if n is large enough, $\bar{X}$ ~ N(μ, σ/√n)

# Review: Normal distribution

- When do we use σ versus σ/√n ?
- We use σ when we want to know the probability of observing a particular value of a normally distributed random variable X
  - E.g. what is the probability of observing a SBP value of 200 when the population mean is 129 and the population SD is 19.8
- We use σ /√n when we are making inference about the sample mean $\bar{X}$

- For normal distribution
  - Stata calculates P(Z<z)
  - If you have a z-value and you want p (PZ<z), use

```
display normal(z)
di normal(1.96)
.9750021
```

  - If you have a z-value and you want p [P(Z>z)], use

```
display 1-normal(z)
di 1-normal(1.96)
.0249979
```

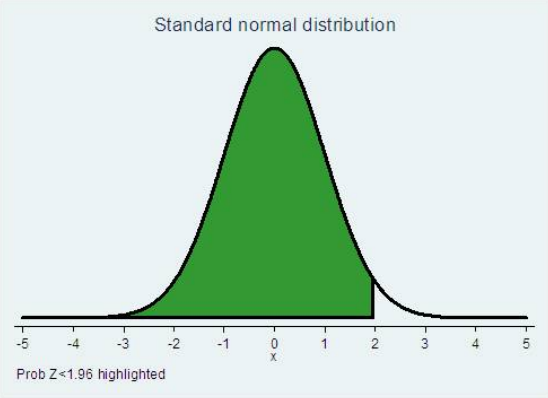
  - If you have p [P(Z<z)] and you want z-value, use

```
display invnormal(p)
. di invnormal(.975)
1.959964
```

  - If you have p [P(Z>z)] and you want z-value, use

```
display invnormal(1-p)
di invnormal(.025)
-1.959964
```

Or use what you know about the symmetry of the distribution



- For t distribution
  - Stata calculates P(T>t) for each d.f.
  - If you have a t value and you want p [P(T>t)], use

```
display ttail(df,t)
di ttail(99,1.96)
.02640356
```

  - If you have a t value and you want p [P(T<t)], use

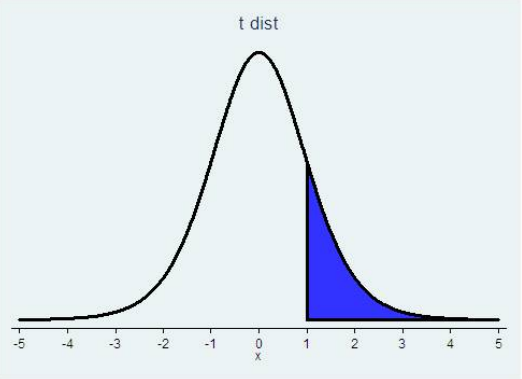
```
display 1-ttail(df,t)
di 1-ttail(99,1.96)
.97359644
```

  - If you have p [P(T>t)] and you want t-value, use

```
display invttail(df,p)
di invttail(99,.025)
1.984217
```

  - If you have p [P(T<t)] and you want t value, use

```
display invttail(df,1-p)
di invttail(99,.975)
-1.984217
```




Or use what you know about the symmetry of the distribution

- Usually you want to summarize the values you observed in your sample and make inference about the sample
- The Central Limit Theorem says that the distribution of a sample mean $\bar{X}$ is a normal distribution with mean μ and standard deviation $\sigma/\sqrt{n}$

# Confidence intervals

- Knowing the distribution of the sample means we write down

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

and do the math to get

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

- If you had drawn your sample 100 times, approximately 95 of the confidence intervals would include the true population mean μ.
  - The interval is random.
- These intervals can be derived for any point estimate, e.g. risk ratio, rate, etc. with known distribution.

- The width of the confidence interval depends on the confidence level (1-$\alpha$) and n
- For confidence intervals for means, if σ is unknown, we use the t distribution and the sample standard deviation s in our computation

$$P(\bar{X} - t_{df,.025}s/\sqrt{n} \leq \mu \leq \bar{X} + t_{df,.025}s/\sqrt{n}) = 0.95$$

# Confidence intervals for proportions

- As n increases, the binomial distribution approaches the normal distribution
- Also, as n increases, the distribution of an estimated proportion approaches the normal distribution
  - $x/n = \hat{p} \sim N( p, \sqrt{(p(1-p)/n)} )$

    x/n is the proportion of successes

    p is the population proportion

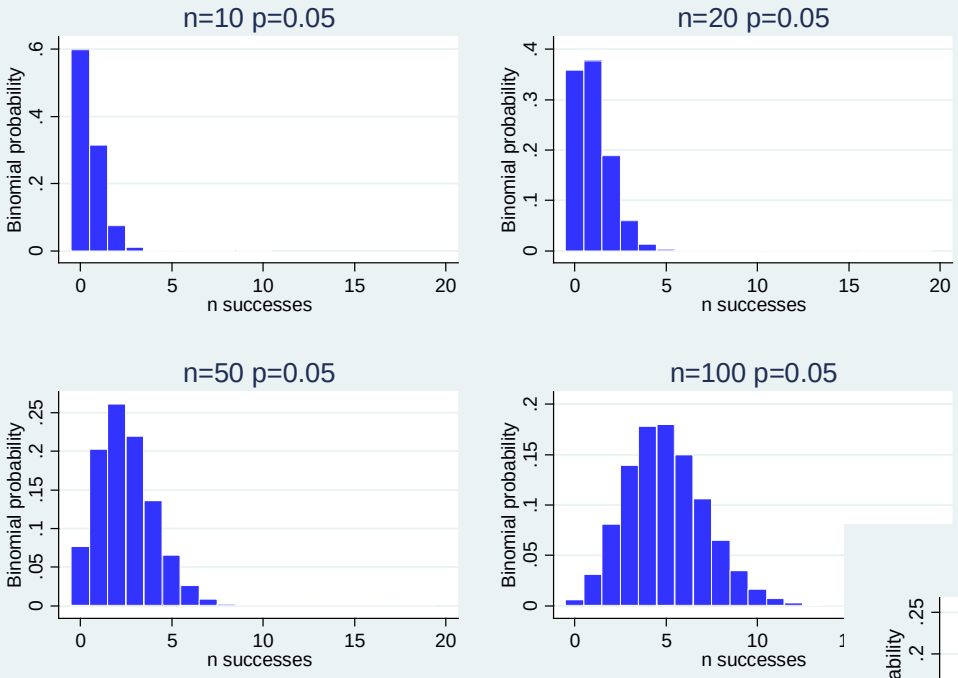
    $\sqrt{(p(1-p)/n)}$ is the standard deviation of x/n if X is binomial

- So we can use the normal distribution to calculate confidence intervals for p

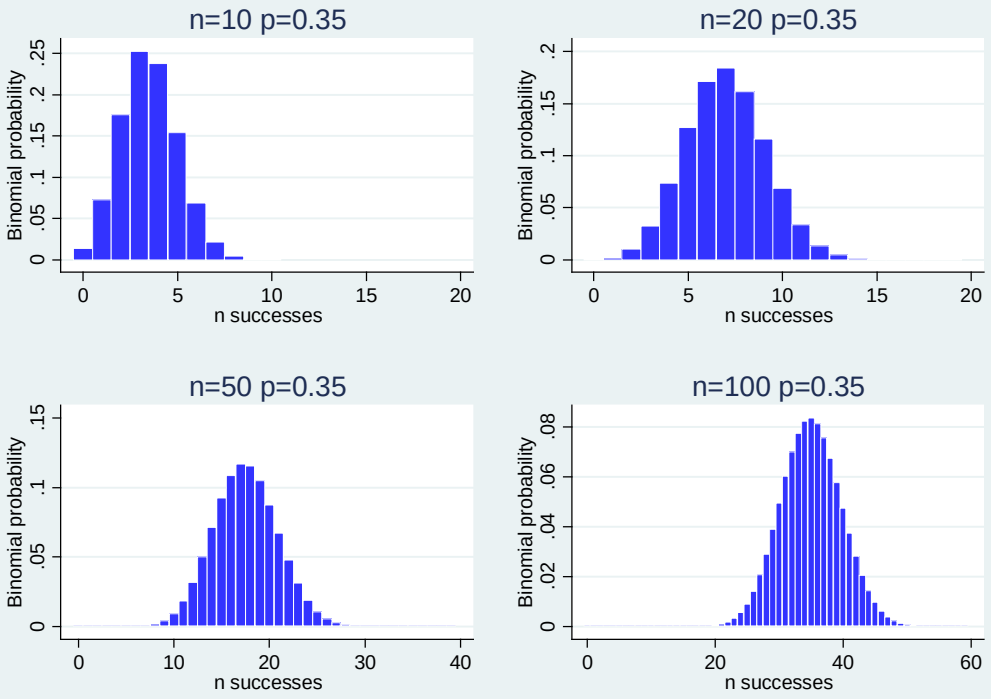
$$\hat{p} - 1.96 * \sqrt{(\hat{p}(1-\hat{p})/n)} , \hat{p} + 1.96 * \sqrt{(\hat{p}(1-\hat{p})/n)}$$

# Confidence intervals for proportions

- But if n is small or p is small or both then the normal approximation is not good and the intervals using the normal approximation are not good
- We need to use the binomial distribution
- Confidence intervals using the binomial distribution are called exact confidence intervals
- Exact confidence intervals are not symmetric around $\hat{p}$ (because neither is the binomial distribution for small n or small p)




For p=0.05, even for large n, the distribution of X doesn't exactly look normal
(100*.05=5)




For p=0.35, the distribution does look fairly normal at small sample sizes. N*.35>5 → n>14.3

# Hypothesis testing

- Confidence intervals tell us about the uncertainty in a point estimate of the population mean or proportion (or some other entity)

- Hypothesis testing allows us to draw a conclusion about a population parameter that we have estimated in a sample

- Remember, we are taking samples from the population and we never know the truth

# Hypothesis testing for a mean

- To conduct a hypothesis test about the mean of a population, we postulate a value for it, and call that value $\mu_0$.
  - We write this $H_0 : \mu = \mu_0$
  - We call this the null hypothesis

- We set an alternative hypothesis for all other values of $\mu$
  - We write this: $H_A : \mu \neq \mu_0$
  - We call this the alternative hypothesis

# Hypothesis testing for a mean

- After formulating the hypotheses, draw a random sample
- Compare the mean of the random sample, $\bar{X}$, to the hypothesized mean $\mu_0$
- Is the difference between the sample mean $\bar{X}$ and the hypothesized population mean $\mu_0$ likely due to chance (remember you have taken a sample), or too large to be attributed to chance?
- If too large to be due to chance, we say we have *statistical significance*

# Hypothesis testing for a mean

- There is some probability that we will incorrectly reject the null hypothesis
  - Incorrectly rejecting the null hypothesis means that we say that the evidence is against the null <u>when the null is really true</u>

# Hypothesis testing for a mean

- We *set* the probability of incorrectly rejecting the null that we are willing to live with in advance, *before collecting the data and doing the analysis*
- That probability is called the *significance level*
- The significance level is denoted by $\alpha$ , and is frequently set to 0.05.

# Hypothesis testing

- Criminal trials – innocent until proven guilty
- We assume the null hypothesis is true, and only reject the null if there is enough evidence that the sample did not come from the hypothesized population (e.g. that the mean is not $\mu_0$)
- In jury trials there is a chance that someone who is innocent is found guilty, but our method of innocent until proven guilty is in place to minimize this

# Hypothesis testing

- In hypothesis testing, a true null hypothesis might be mistakenly rejected (if the sample you drew very different from the null just by chance)
- This mistake is called a Type I error
  - The probability of a Type I error is chosen in advance
  - This probability is called the significance level, $\alpha$

# p value for testing a mean

- The probability of obtaining a mean as or more extreme as the observed sample mean $\bar{X}$, *given that the null hypothesis is true,* is called the *p-value* of the test, or p.

# Basic steps

- Before the test, set the significance level, $\alpha$
  - This is the probability of rejecting the null hypothesis when it is true P(Type I error) that you can live with
- When you run the test, you get a probability of observing a sample mean as extreme as you did, given that the null hypothesis is true
  - This probability is the p-value
- If the p-value is less than $\alpha$, (e.g. $\alpha$=0.05 and p=0.013), then the test is called statistically significant and the null hypothesis is rejected

# Hypothesis testing steps

- Specify the null and alternative hypothesis
- Determine the compatibility of the data with the null hypothesis
- Either:
  - Reject the null

    OR
  - Fail to reject the null.  NEVER accept the null.

# Hypothesis testing

- Specify the null and alternative hypothesis
  - E.g. Null: Mean CD4 cell count in population is ≥350
    Alt: " < 350
- Determine the compatibility of the data with the null hypothesis
  - Calculate a test statistic and compare it to its corresponding theoretical probability distribution to get the p-value
  - If the test statistic is very large, then the data are very unlikely under the null hypothesis
- Either:
  - Reject the null if the p-value < $\alpha$ OR
  - Fail to reject the null

# 1. State the hypotheses

- In epidemiology the null hypothesis is often that there is no association between the exposure and the outcome

  E.g.,

  - The difference in disease risk=0
    - $R_{exposed} - R_{not\ exposed} = 0$
  - The ratio of risk/prevalence/etc=1
    - $R_{exposed} / R_{not\ exposed} = 1$
  - Clinical trials: The mean value of something does not differ between two groups
- The alternative hypothesis is usually an association or risk difference
  - E.g.,
    - $R_{exposed} - R_{not\ exposed} \neq 0$
    - $R_{exposed} / R_{not}$ exposed $\neq 1$
    - The mean response in the treated group $\neq$ the mean response in the untreated group

# 1. State the hypotheses

- $H_0$ (the null) and $H_A$ (the alternative) must be mutually exclusive (no overlap) and include all the possibilities
- The alternative is the complement of the null

# 2. Determine the compatibility with the null

- We determine if there is enough evidence to reject the null hypothesis (i.e. calculate the test statistic and find the p-value)
  - If the null is true, what is the probability of obtaining the sample data as extreme or more extreme?
    - Calculate a test statistic
    - Find the probability of the test statistic if the null is true
  - This probability is call the p-value

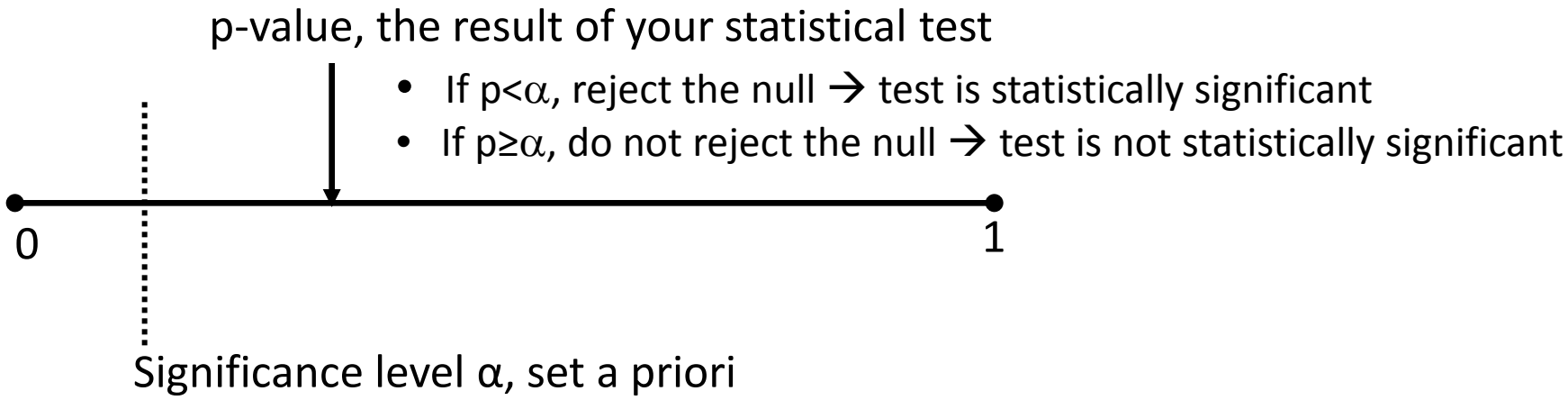
# 3. Reject or fail to reject

- If the p-value (the probability of obtaining the sample data and corresponding test statistic if the null is true) is sufficiently small (we often use 5%) then we reject the null and say the test was statistically significant.
- This 5% is $\alpha$, the significance level

# 3. Reject or fail to reject

- If we fail to reject the null hypothesis, it does not mean that we accept it.
  - We might fail to reject the null if the sample size is too small or the variability in the sample was too large
  - We say we "reject the null" (if the p-value is $< \alpha$) or we "fail to reject the null" (if the p-value is $\geq \alpha$)

# Hypothesis testing



# Tests of one mean

- We want to test whether a mean is equal to some hypothesized value

- If we believe there might be deviations in either direction, we use a two-sided test
- Two sided test:
  - Null hypothesis: $H_0$: $\mu=\mu_0$
  - Alternative hypothesis: $H_A$: $\mu\neq\mu_0$

# Tests of one mean

- We want to test whether a mean is equal to some hypothesized value

- If we only care about values above or below a certain value, we use a one-sided test
- One sided test:
  - Null hypothesis: $H_0$: $\mu \geq \mu_0$
  - Alternative hypothesis: $H_A$: $\mu < \mu_0$

or

  - Null hypothesis: $H_0$: $\mu \leq \mu_0$
  - Alternative hypothesis: $H_A$: $\mu > \mu_0$

# Lexicon

- For two sided tests, people often say they are testing the hypothesis that the mean is $\mu_0$ . When they say this they are stating the null hypothesis.
- Since you know that the null is the complement of the alternative, you don't usually state both in practicality (but we will for this class)

# Test of one mean

- The distribution of the sample mean $\bar{x}$ if n is large enough is normal by the CLT. So you can calculate the z statistic:

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

If σ is not known, calculate

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

# Tests of one mean

- Compare the test statistic to the appropriate distribution to get the p-value

- For a one-sided test, if the alternative hypothesis is $H_A: \mu > \mu_0$ then the p-value is

$$P(Z > z_{stat}) \text{ or } P(T > t_{stat})$$

- For a one-sided test, if the alternative hypothesis is $H_A: \mu < \mu_0$ then the p-value is

$$P(Z < z_{stat}) \text{ or } P(T < t_{stat})$$

# Tests of one mean

- For a two-sided test, the p-value is

$$P(Z>z_{stat})+P(Z<-z_{stat}) = 2*P(Z>z_{stat})$$

OR $$P(T>t_{stat})+P(T<-t_{stat}) = 2*P(T>t_{stat})$$

# Test of one mean, one sided

Non-pneumatic anti-shock garment for the treatment of obstetric hemorrhage in Nigeria

- Mean initial blood loss 1413.1 ml; SD=491.3; n=83
- Our question: Are these women sampled from a population that is hemorrhaging (>750 ml blood loss)?

Data adapted from Miller, S et al. Int J Gynecol Obstet (2009)

# Test of one mean, one sided

- The null hypothesis is they are not hemorrhaging

$H_0$: $\mu \leq 750$  $H_A$: $\mu > 750$  $\alpha = 0.05$

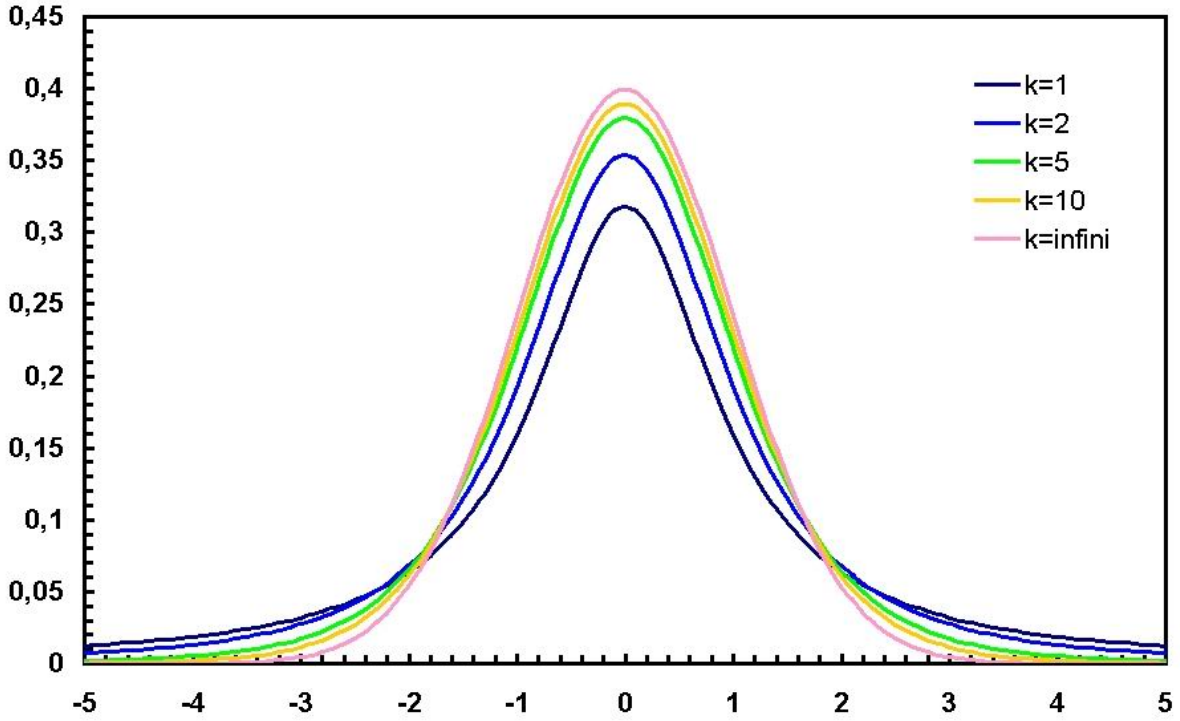
- T-statistic: $t_{stat} = \dfrac{\overline{X} - \mu_0}{s/\sqrt{n}}$

$$= \frac{1413.1 - 750}{491.3/\sqrt{83}} = 12.3$$

- p-value: P(T>12.3) with 82 d.f.

Data adapted from Miller, S et al. Int J Gynecol Obstet (2009)

# Test of one mean



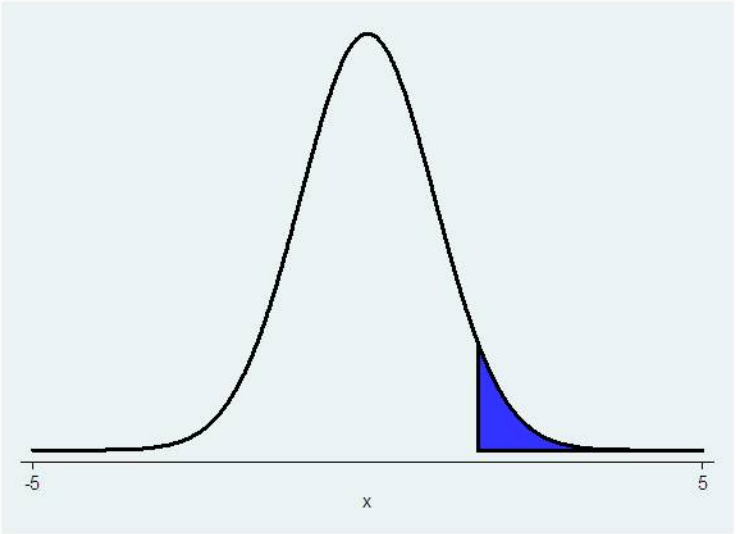

- 12.3 is off the graph

- So the probability of observing a sample mean of 1413 with n=83 and SD=431 if the true mean is <=750 is very very low

# Test of one mean, one sided hypothesis, Stata

P(T>test stat)

```
. display ttail(82, 12.3)
1.308e-20
```




The p-value is <0.001, which is a lot less than 0.05, therefore we reject the null

# Test of one mean, one sided hypothesis, Stata

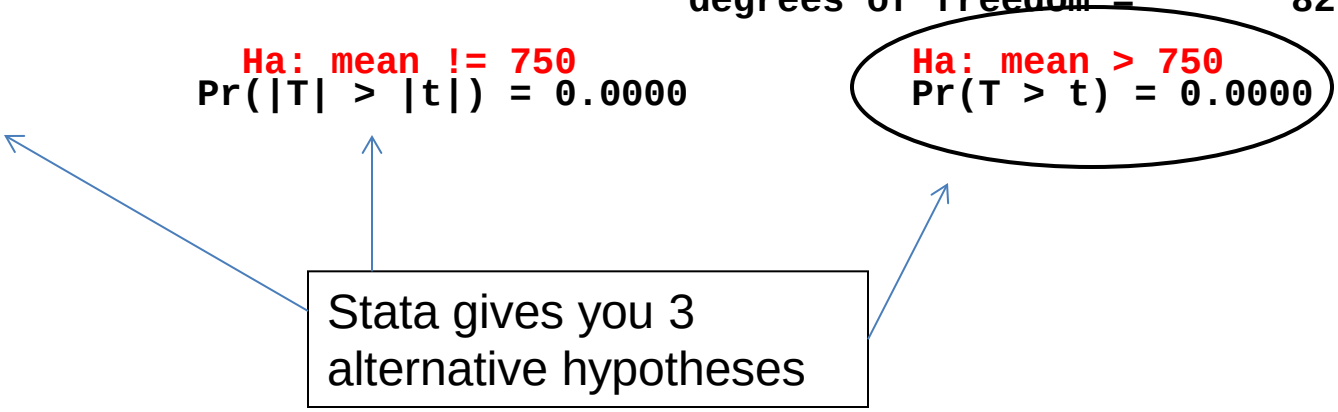
Another way: Stata *immediate* code for ttests:

ttesti samplesize samplemean samplesd hypothesizedmean

```
ttesti 83          1413.1       491.3        750

One-sample t test
------------------------------------------------------------------------------
         |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       x |      83      1413.1    53.92718       491.3    1305.822    1520.378
------------------------------------------------------------------------------
    mean = mean(x)                                                t =  12.2962
    Ho: mean = 750                                 degrees of freedom =       82

   Ha: mean < 750               Ha: mean != 750                 Ha: mean > 750
   Pr(T < t) = 1.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 0.0000
```




→ We reject the null

# Test of one mean, two sided hypothesis

Example: Do children with congenital heart disease walk at the same or a different age as compared to healthy children? Healthy children start walking at mean age of 11.4 months

- Null hypothesis: $H_0$: $\mu$= 11.4 months ($\mu_0$)
- Alternative hypothesis: $H_A$: $\mu \neq$ 11.4 months
- Significance level=0.05
- Data: Sample of children with congenital heart defects, n=9, sample mean=12.8, sample SD=2.4

- Calculate the test statistic and its associated p value

$$t_{stat} = \frac{\overline{X} - \mu_0}{s/\sqrt{n}}$$

$$= \frac{12.8 - 11.4}{2.4/\sqrt{9}} = 1.75$$

- p- value is P(T>1.75) + P(T<-1.75) = 2*P(T>1.75)

```
display ttail(8,1.75)*2
.11823278
```

Or run ttesti in Stata

** ttesti  samplesize  samplemean  samplesd  hypothesizedmean

```
ttesti 9 12.8 2.4 11.4

One-sample t test
------------------------------------------------------------------------------
         |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       x |       9        12.8          .8         2.4     10.9552     14.6448
------------------------------------------------------------------------------
    mean = mean(x)                                                t =   1.7500
Ho: mean = 11.4                                 degrees of freedom =        8

   Ha: mean < 11.4               Ha: mean != 11.4                 Ha: mean > 11.4
 Pr(T < t) = 0.9409         Pr(|T| > |t|) = 0.1182          Pr(T > t) = 0.0591
```

→ Fail to reject the null

# Test of one mean

- For data already in Stata, the ttest command also works
- E.g. We want to test that the mean CD4 cell count <350 among persons newly diagnosed with HIV in Uganda
  - Null hypothesis: $H_0$: $\mu \geq 350$ cells/$mm^3$
  - Alternative hypothesis (the one we are worried about – that patients come in with low CD4 cell counts):

    $H_A$: $\mu < 350$ cells/$mm^3$
  - Significance level=0.05

# Test of one mean

## Use ttest varname== hypothesized value

- **ttest cd4count==350**

```
One-sample t test
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
cd4count |     999    329.2332    8.419592    266.1177    312.7111    345.7554
------------------------------------------------------------------------------
    mean = mean(cd4count)                                         t =  -2.4665
Ho: mean = 350                                 degrees of freedom =      998

    Ha: mean < 350               Ha: mean != 350                 Ha: mean > 350
 Pr(T < t) = 0.0069         Pr(|T| > |t|) = 0.0138          Pr(T > t) = 0.9931
```

→ We reject the null

Data set: vct_baseline_biostat200_v1.dta

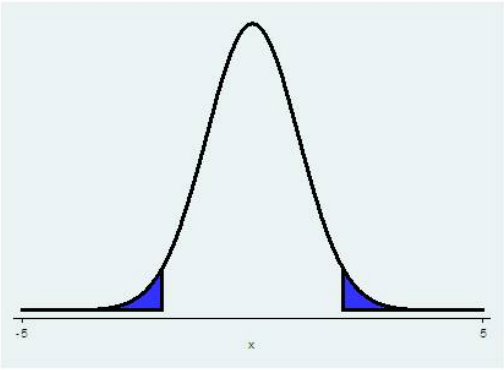
# Hypothesis testing

- One sided hypothesis : e.g. $H_A$: $\mu > \mu_0$
  - For a one sided hypothesis, if you know $\sigma$, you calculate $P(Z>z)$
  - You are only looking at one tail of the distribution
  - For $P(Z>z)$ to be $<0.05$, the test statistic $z_{stat}$ must be $>1.645$

# Hypothesis testing

- Two sided hypotheses $H_A$: $\mu \neq \mu_0$
  - For a two sided hypothesis, you are considering the probability that $\mu > \mu_0$ or $\mu < \mu_0$ , so you calculate

    P(Z>z) or P(Z<-z) if you know σ, which is both tails of the distribution
  - This is then 2*P(Z>z)
  - For 2*P(Z>z) to be <0.05, the area in each tail must be <0.025, so the test statistic $z_{stat}$ must be >1.96 or <-1.96



# Hypothesis testing

- So at the same significance level, here $\alpha$=0.05, a smaller test statistic, and therefore less evidence, is needed to reject the null for a one sided test as compared to a two sided test

- So a two sided test is more conservative

# Hypothesis testing

- What if you ran a test and got $z_{stat}$ =1.83? If your hypothesis was one-sided, you'd reject. If your hypothesis was two-sided, you'd fail to reject.

- Therefore it is very important to specify your test a priori!

# Hypothesis testing

- For clinical trials, most people use two-sided hypotheses
  - That way no one will suspect you of changing your hypothesis test after the fact
  - On the other hand, does it really make sense to have an alternative hypothesis that a new treatment is either better *or worse* than the old one?

# Inference on proportions

- Variables that are measured as successes or failures, disease or no disease, etc., can be considered binomial random variables
  - x=number of successes
  - n=number of “trials”
  - x/n = proportion diseased

# Hypothesis test of a proportion

- Under the CLT, the sampling distribution of an estimated proportion $\hat{p}$, if n is large enough, is approximately a normal distribution with mean p and standard deviation=$\sqrt{(p(1-p)/n)}$

# Hypothesis test of a proportion

- Therefore we can test that a sample proportion is equal to, greater than, or less than some hypothesized $p_0$ using the z statistic

$$z_{stat} = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$$

# Hypothesis test of a proportion

- Example:  Micronutrient intake of black women in South Africa
- Study:  Pre-menopausal women randomly selected based on geographic location
- Micronutrient intake was determined using a Quantitative Food Frequency Questionnaire developed for South Africans
- Question:  Do more than 10% of the women have the micronutrient intake of less than the RDA?

Data adapted from Hattingh Z. et al, West Indian Med J 2008: 57 (5):431

# Hypothesis test of a proportion

- Null hypothesis:
  - 10% or fewer of the women aged 25-34 have insufficient dietary folate levels (less than 268 μg, a cutoff based on RDA)
  - $H_0$: $p_0 \leq 0.10$
- Alternative hypothesis:
  - More than 10% of the women aged 25-34 have insufficient dietary folate levels
  - $H_A$: $p_0 > 0.10$
- Significance level set at = 0.05

# Hypothesis test of a proportion

- The data:
  - 158/279=56.6% had folate levels<268 μg
- Using the normal approximation to the binomial distribution: $z_{stat} = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$

$z_{stat}$ = (.566-0.10)/√(0.10*0.90/279) =25.9

The chance of observing a z statistic of this large (25.9) is very small (<0.05)

→ So we reject the null hypothesis

# Hypothesis testing of a proportion

The stata command prtest or prtesti gives you a test of a proportion using the normal approximation

** prtesti samplesize observedp hypothp

```
. prtesti  279 .566  .1

One-sample test of proportion                      x: Number of obs =      279
------------------------------------------------------------------------------
    Variable |       Mean   Std. Err.                     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |       .566   .0296723                      .5078434    .6241566
------------------------------------------------------------------------------
    p = proportion(x)                                             z =  25.9458
Ho: p = 0.1

     Ha: p < 0.1                 Ha: p != 0.1                  Ha: p > 0.1
 Pr(Z < z) = 1.0000         Pr(|Z| > |z|) = 0.0000          Pr(Z > z) = 0.0000
```

# Using the binomial distribution

- This is a straightforward test of the binomial distribution -- we are asking what is the probability that we observe 158 events when n=279 and the "true" proportion is 0.10

```
. di binomialtail(279,158,.1)
 1.268e-82
```

# Using the binomial distribution

The Stata command bitest or bitesti also does this

```
**   bitesti samplesize observed_proportion hypoth_proportion

. bitesti 279 .566 .1

         N   Observed k   Expected k   Assumed p   Observed p
------------------------------------------------------------
       279         158         27.9       0.10000      0.56631

  Pr(k >= 158) = 0.000000  (one-sided test)
  Pr(k <= 158) = 1.000000  (one-sided test)
  Pr(k >= 158) = 0.000000  (two-sided test)

  note: lower tail of two-sided p-value is empty
```

Use i at the end (ttesti prtesti bitesti) when you do not have the data in stata but want to run a test

# Hypothesis testing of a proportion example

- Null hypothesis:
  - Fewer than 10% of the women aged 25-34 have dietary zinc levels of less than 5 mg
  - $H_0$: $p < 0.10$
- Alternative hypothesis:
  - 10% or more of the women aged 25-34 have dietary zinc levels of less than 5 mg
  - $H_A$: $p \geq 0.10$
- Significance level = 0.05

# Hypothesis testing of a proportion

- The data: 31/279=11.1% had zinc levels of less than 5 mg

```
prtesti 279 .111 .1

One-sample test of proportion                      x: Number of obs =      279
------------------------------------------------------------------------------
    Variable |       Mean   Std. Err.                      [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |       .111   .0188066                      .0741397    .1478603
------------------------------------------------------------------------------
    p = proportion(x)                                             z =   0.6125
Ho: p = 0.1

     Ha: p < 0.1                 Ha: p != 0.1                  Ha: p > 0.1
 Pr(Z < z) = 0.7299         Pr(|Z| > |z|) = 0.5402          Pr(Z > z) = 0.2701
```

- Using the normal approximation, we find that the data are NOT inconsistent with the null hypothesis, therefore we fail to reject the null

# Using the binomial distribution

```
. bitesti 279 .111 .1

        N   Observed k   Expected k   Assumed p   Observed p
------------------------------------------------------------
      279          31         27.9       0.10000      0.11111

  Pr(k >= 31)            = 0.295225  (one-sided test)
  Pr(k <= 31)            = 0.767742  (one-sided test)
  Pr(k <= 24 or k >= 31) = 0.548577  (two-sided test)
```

```
Or we could have used:
. di binomialtail(279,31,.1)
.29522463
```

These are exact tests, rather than using the normal approximation.

# Hypothesis testing of a proportion

- Remember that if n or p are small, you may need to use an exact test
- The commands in Stata to do this are

```
display binomialtail(n,k,p)
```

or

```
bitesti samplesize  observedp   hypothp
```

# Hypothesis tests versus confidence intervals

- You can reach the same conclusion with confidence intervals as with two-sided hypothesis tests
  - Reject the null if the 95% confidence interval does not include the hypothesized value $\mu_0$
- Confidence intervals give us more information: the range of reasonable values for μ and the uncertainty in our estimate $\bar{X}$ .
- Many statisticians and others prefer using confidence intervals to using hypothesis tests.

# Types of error

- Type I error – significance level of the test

  $\alpha$=P(reject $H_0$ | $H_0$ is true)
- Occurs when we incorrectly reject the null
- We take a sample from the population, calculate the statistics, and make inference about the true population. If we did this repeatedly, we would incorrectly reject the null 5% of the time *that it is true* if $\alpha$ is set to 0.05.

# Types of error

- Type II error –

  $\beta$ = P(do not reject $H_0$ | $H_0$ is false)

- Occurs when we incorrectly fail to reject the null, i.e. when we do not reject the null *when the null is false*

# Types of error

- Remember, $H_0$ is a statement about the population and is either true or false, but we don't know which
- We take a sample and use the information in the sample to try to determine the answer
- We set our chance of a Type I error, and can design our study to minimize the chance of a Type II error

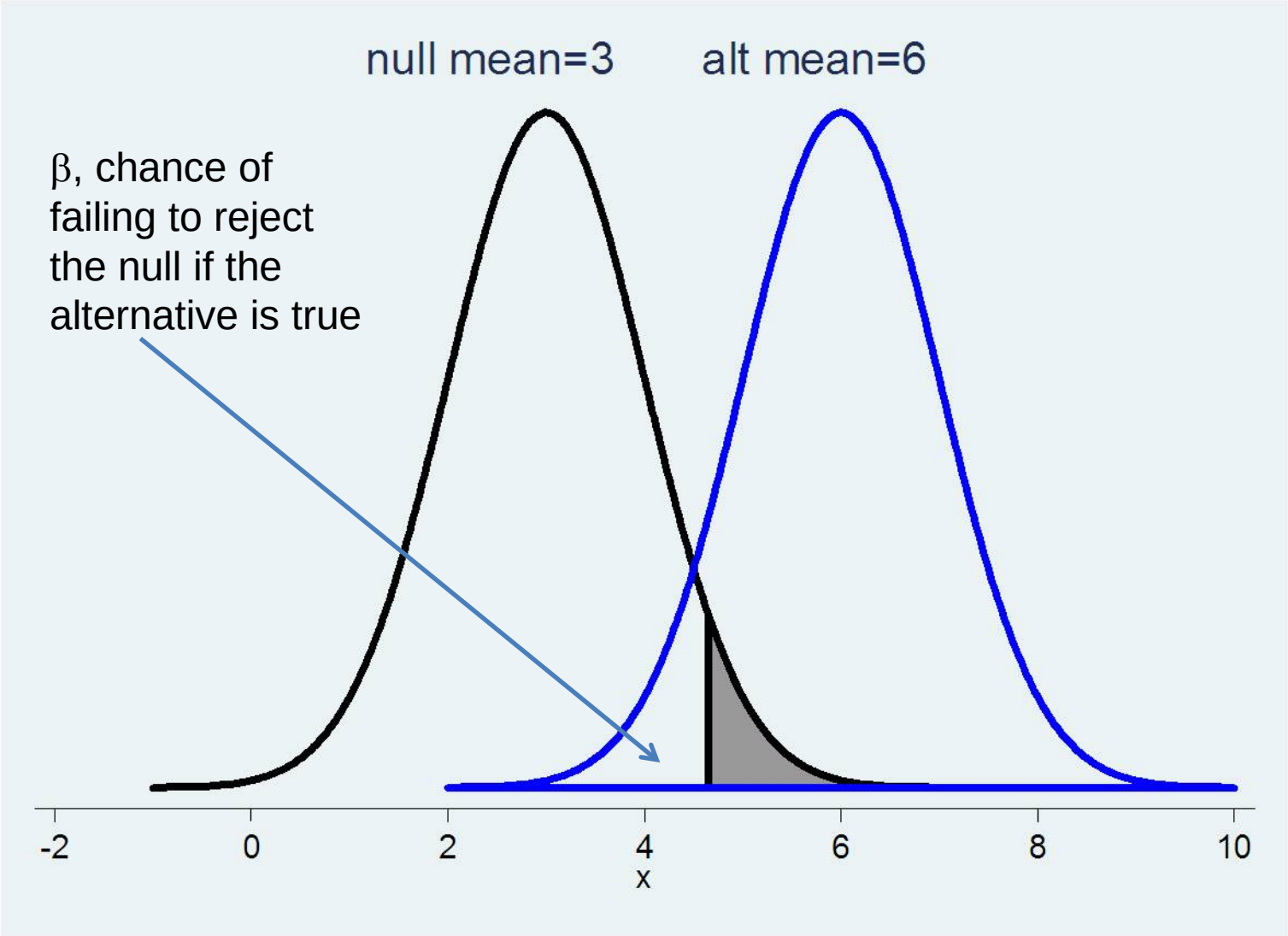
| | True state | |
|---|---|---|
| **Decision** | **$H_0$ is true** | **$H_0$ is false** |
| Do not reject $H_0$ | Correct | Type II error=$\beta$ |
| Reject $H_0$ | Type I error=$\alpha$ | Correct |

# Type II error

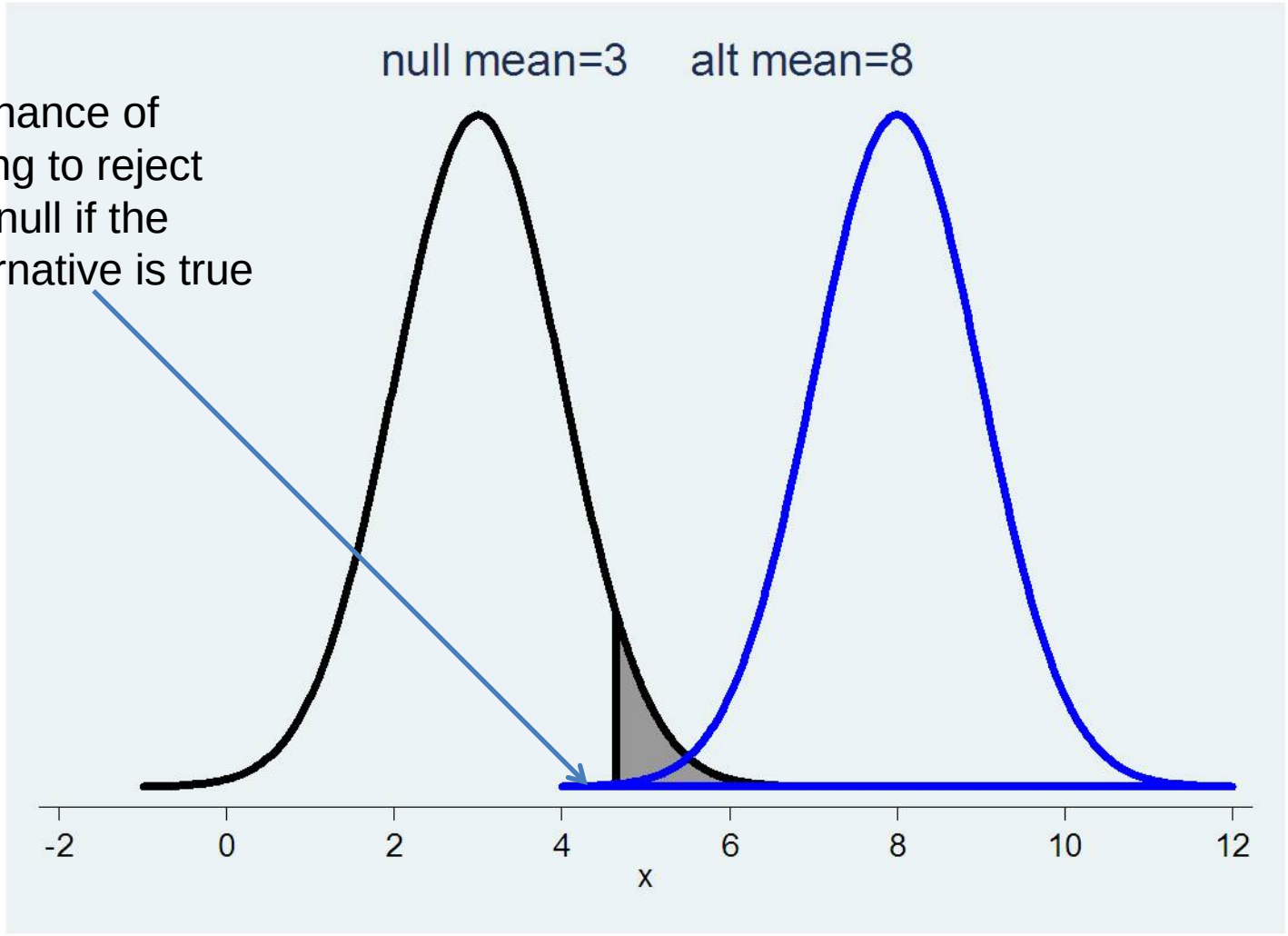
- When you miss the effect (statistically)
- This will happen when your test statistic is not high enough
- For a mean, this would be

$$t_{stat} = \frac{(\bar{X} - \mu_0)}{s/\sqrt{n}}$$

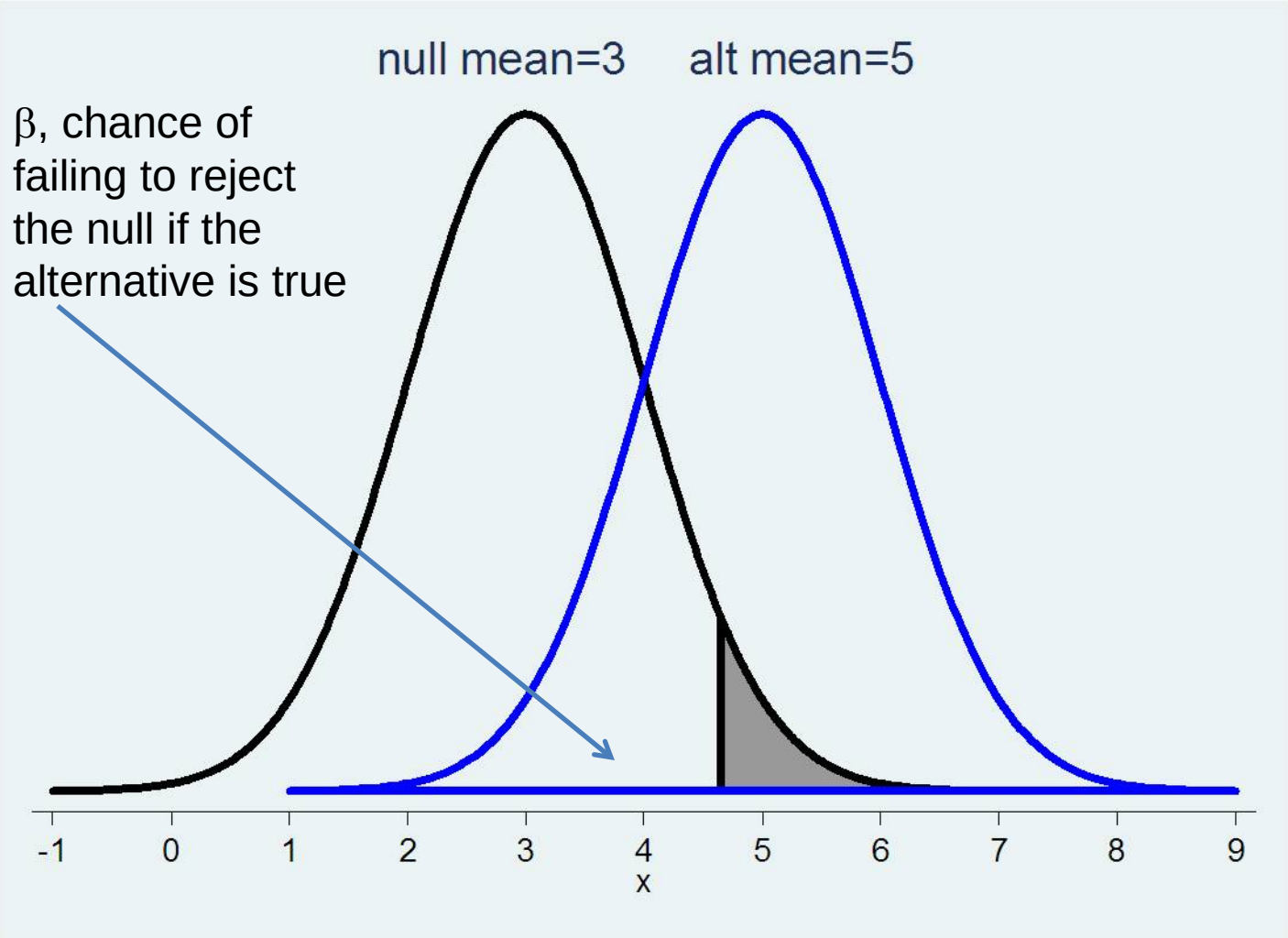
# Chance of a type II error



# If the alternative is very different from the null, the chance of a Type II error is low



# If the alternative is not very different from the null, the chance of a Type II error is high



# Finding β, P(Type II error)

- Example: Mean age of walking
  - $H_0$: μ< 11.4 months ($\mu_0$) $H_A$: μ>11.4 months
  - Known SD=2
  - Sample size=9

- We will reject the null if the $z_{stat}$ (assuming σ known) > 1.645
- So we will reject the null if $z_{stat} \geq \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} \geq 1.645$

$$\bar{x} \geq 1.645\sigma / \sqrt{n} + \mu_0$$

- For our example, the null will be rejected if

$\bar{X}$> 1.645*2/3 + 11.4 > 12.5

- If the true mean $\mu$ is really 14 (versus the null value of <11.4, meaning the null is false), what is the probability that the null will not be rejected?
  - The probability of a Type II error?
- The null will be rejected if the sample mean is >12.5, not rejected if it is ≤12.5
- What is the probability of getting a sample mean of ≤12.5 if the true mean is 14? This is a z statistic.
- P(Z<(12.5-14)/(2/3))

```
di normal((12.5-14)/(2/3))
.01222447
```

So if the true mean is 14 and the sample size is 9, the probability of incorrectly failing to reject the null is 0.012

- Note that this depended on this other population mean (e.g. 14)
  - The chance of failing to reject the null will increase if this true population mean is closer to the null value
- What is the probability of failing to reject the null if the true population mean is 13?
  - P(Z<(12.5-13)/(2/3))

```
di normal((12.5-13)/(2/3))
.22662735
```

- So for an alternative mean that is closer to the null value, the chances of failing to reject are higher

# Power

- 1-$\beta$ is called the power of a statistical test
- The power of a test is the probability that you will reject the null hypothesis when you should (i.e. when it is false)
- You can construct a power curve by plotting power versus different alternative hypotheses
- You can also construct a power curve by plotting power versus different sample sizes (n's)
  - This curve will allow you to see what gains you might have versus cost of adding participants
  - Power curves are not linear

# Power

- The power of a statistical test is lower for alternative values that are closer to the null value.
    - When alt mean was 14, power =1-.012=.988
    - When alt mean was 13, power was 1-.227=.773
- The power of a statistical test can also be increased by increasing n
- For a fixed n, the power of a statistical test can be increased by increasing $\alpha$, the probability of a Type I error

# Power

- The balance between type I error and type II error (or power) should depend on the context of the study
- When setting up a study, most investigators make sure that there will be at least 80% power.
- Sample size formulas allow you to specify the desired $\alpha$ and $\beta$ levels.

# For next time

- Read Pagano and Gauvreau
  - Chapter 10 and 14 (pages 329-330) today's material
  - Pagano and Gauvreau Chapters 11-12, and 14 (pages 332-338)