Today

- Where are we & where are we going
- Brief Review of finding percentiles and cutoff values, CLT and 95% confidence intervals
- Hypothesis testing in general
- Hypothesis testing of means
- Hypothesis testing of proportions
- Types of error

Where are we & what's next?

- Week 1: Variables, tables, graphs
- Week 2: Probability: We have rules!
  - Independence → multiplicative rule
  - Mutual exclusivity → additive rule
  - Conditional probability → Bayes rule
- 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 means or proportions and error types
- Weeks 6-7: Multiple comparisons, non-parametric tests, & more
- Week 8: Visualization
- Weeks 9-11: Correlation, regression, & logistic regression & more

- 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
```

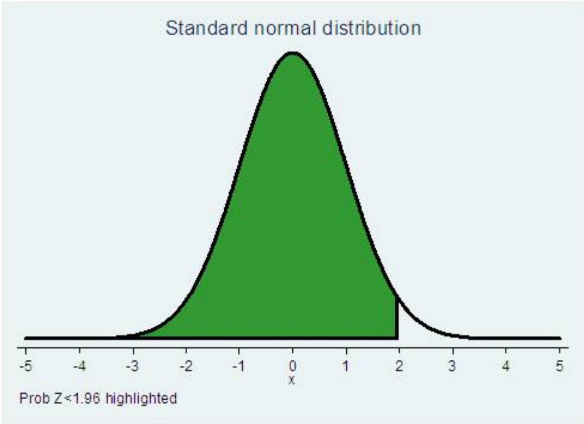
  - If you have p [P(Z<z)] and you want z-value, use

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

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

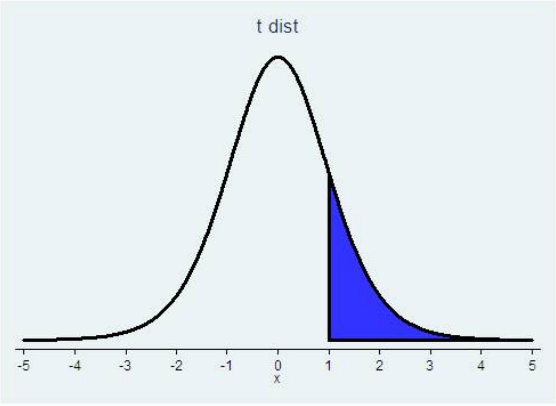
```
display invnormal(1-p)
di invnormal(.025)
-1.959964
```

    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

Confidence intervals

- Knowing the distribution of the sample means we write down

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

and do the math to rearrange

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

- 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.
- We use either z or t depending on whether the variance is known

Confidence intervals for <u>proportions</u>

- As n increases, the binomial distribution approaches the normal distribution
- Also, as n increases, the distribution of an estimated <u>proportion</u> 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 <u>proportions</u>

- 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 the binomial distribution is not symmetric for small n or small p)

Hypothesis testing

- Confidence intervals tell us about the uncertainty in a point estimate of the population mean or proportion

- 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.

Statistical hypothesis tests: cheat sheet

Data and comparison type	Alternative hypotheses	Parametric test Stata command
Numerical; One mean	H_a: $\mu \neq \mu_a$ (two-sided) H_a: $\mu > \mu_a$ or $\mu < \mu_a$ (one-sided)	Z or t-test •ttest var1==hypoth val.*
Numerical; Two means, paired data	H_a: $\mu_1 \neq \mu_2$ (two-sided) H_a: $\mu_1 > \mu_2$ or $\mu < \mu_a$ (one-sided)	Paired t-test •ttest var1==var2*
Numerical; Two means, independent data	H_a: $\mu_1 \neq \mu_2$ (two-sided) H_a: $\mu_1 > \mu_2$ or $\mu < \mu_a$ (one-sided)	T-test (equal or unequal variance) •ttest var1, by(byvar) •ttest var1, by(byvar) unequal
Numerical, Two or more means, independent data		
Dichotomous; One proportion	H_a: $p \neq p_a$ (two-sided) H_a: $p > p_a$ or $p < p_a$ (one-sided)	Proportion test • prtest var1=hypoth value* • bitest var1=hypoth value
Dichotomous; two proportions	H_a: $p_1 \neq p_2$ (two-sided) H_a: $p_1 > p_2$ (one-sided)	Proportion test (z-test) • prtest var1, by(byvar)

Hypothesis testing for a <u>mean</u>

- To conduct a hypothesis test about the mean of a population, we hypothesize a value for it, and call that value μ_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 μ
  - We write this: $H_A : \mu \neq \mu_0$
  - We call this the alternative hypothesis

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 α or alpha, and is frequently set to 0.05 or 5% error

Hypothesis testing

- In hypothesis testing, a true null hypothesis might be mistakenly rejected (if the sample you drew was 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, α

The probability of obtaining a mean at or more extreme than the observed mean, *given that the null hypothesis is true,* is the p-value of the test.

Basic steps in Hypothesis testing

- Before the test, set the significance level, α
  - 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 α, (e.g. α=0.05 and p=0.013), then the test is called statistically significant and the null hypothesis is rejected

Hypothesis Testing
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

Hypothesis Testing
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

Hypothesis Testing
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 α, the significance level

- <u>If we fail to reject the null hypothesis, it does not mean that we accept it.</u>

Hypothesis Testing of one mean

- We want to test whether a mean is equal to an 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: $\mu\neq\mu_0$
- 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: $\mu>\mu_0$

Hypothesis Testing 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

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

Hypothesis Testing 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)

Hypothesis Testing of one mean: one-sided

- The null hypothesis is they are not hemorrhaging

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

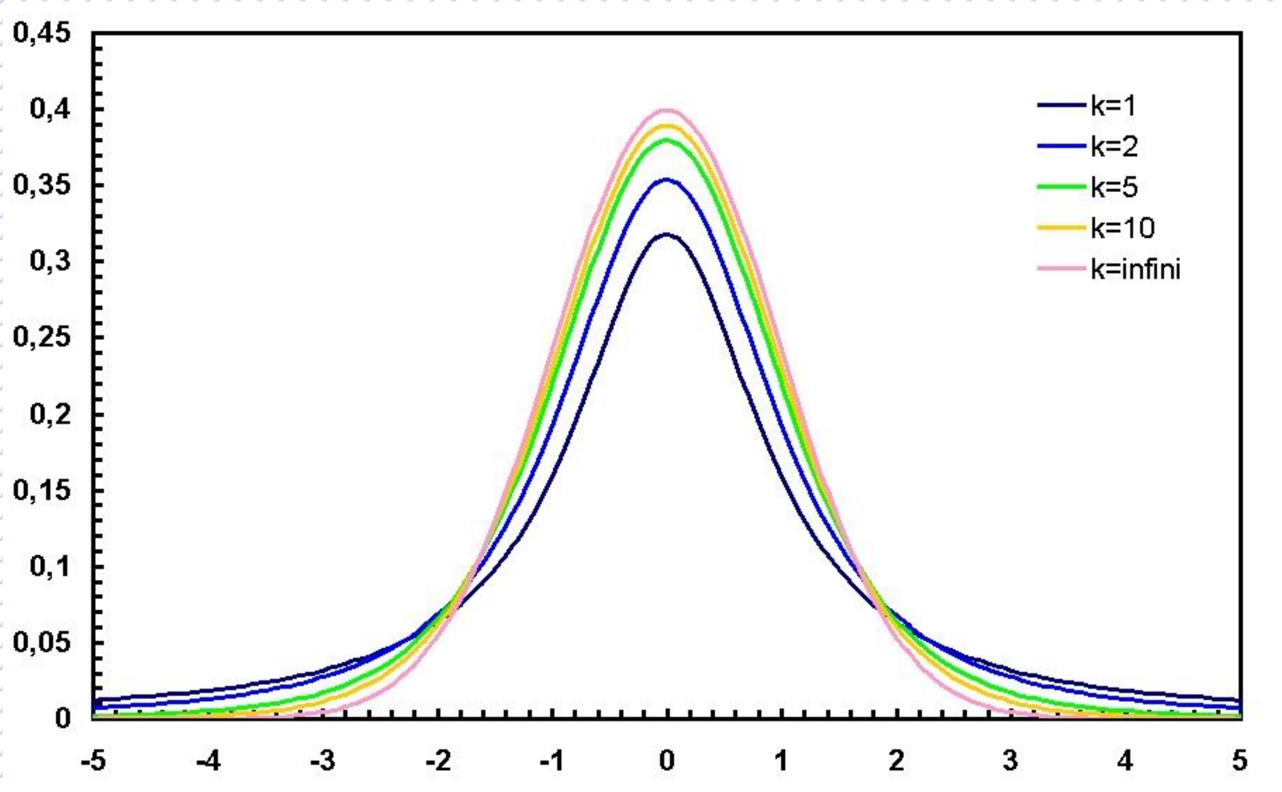
- T-statistic: $$t_{stat} = \frac{\overline{X} - \mu_0}{s / \sqrt{n}}$$

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

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

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

Hypothesis Testing of one mean



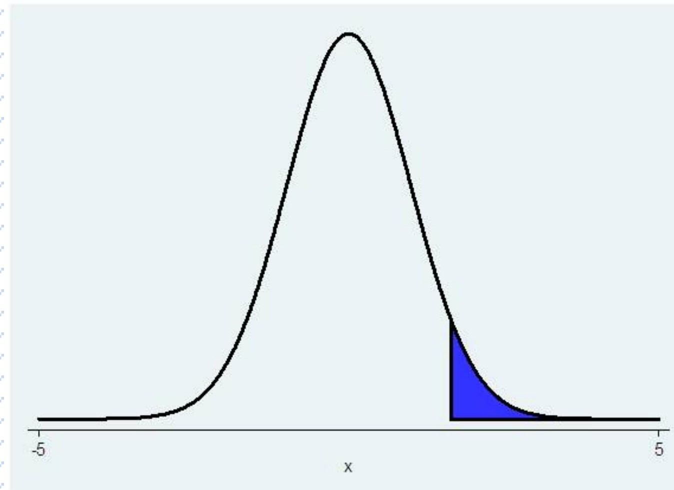
- 12.3 is off the graph

- 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

Hypothesis Testing of one mean: one-sided

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

Hypothesis Testing of one mean: one-sided

Another way: Stata *calculate* 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
```

Stata gives you 3 alternative hypotheses

→ We reject the null hypothesis

Hypothesis Testing, 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 (μ_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

Hypothesis Testing, two sided hypothesis

- 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
```

Hypothesis Testing, two sided hypothesis

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

Hypothesis Testing, one or two sided hypotheses

- 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):

    $$H_A: \mu < 350 \text{ cells/mm}^3$$

  - Significance level=0.05

Hypothesis Testing, one or two sided hypothesis

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 or two sided hypothesis

- 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 <u>specify your test a priori</u>!

Hypothesis testing

- For clinical trials, most of the time we use a 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?
  - Assuming your new therapy is better or ‘not inferior’ we may use a one-sided test

Hypothesis test of a proportion

- 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
  - Use 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
```

Hypothesis testing of a proportion

- 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
```

This is an exact test, 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 – we will examine other tests (Chi-squared) in future lectures
- The commands in Stata to do this are

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

or

```
bitesti samplesize  observedp   hypothp
```

Comparison of two means: the *paired* t-test

- Paired samples, numerical variables
  - Two determinations on the same person (before and after) – e.g. before and after intervention
  - Matched samples – measurement on pairs of persons similar in some characteristics, i.e. identical twins (matching is on genetics)
- Each person or pair has 2 data points, and we calculate the difference for each
- Then we can use our one-sample methods to test hypotheses about the value of the difference

Comparison of two means: *paired* t-test

- Step 1: The hypotheses (two sided)
- Generically H_0: $\mu_1 - \mu_2 = \delta$

  H_A: $\mu_1 - \mu_2 \neq \delta$

- Often $\delta = 0$, no difference

So H_0: $\mu_1 - \mu_2 = 0$, i.e. H_0: $\mu_1 = \mu_2$

H_A: $\mu_1 - \mu_2 \neq 0$, i.e. H_A: $\mu_1 \neq \mu_2$

Comparison of two means: *paired* t-test

- Step 2: Calculate the test statistic

$$t_{stat} = \frac{\bar{d} - \delta}{s_d / \sqrt{n}} \text{ where}$$

$\bar{d}$ is the average of the differences d_i between the pairs

$$s_d = \sqrt{\frac{\sum_{i=1}^{n} (d_i - \bar{d})^2}{n-1}}$$

δ is the hypothesized difference between the means

Comparison of two means: *paired* t-test

- Step 3: Reject or fail to reject the null
  - Is the p-value (the probability of observing a difference as large or larger, under the null hypothesis) greater than or less than the significance level, α?

Comparison of two means: *paired* t-test

Example: BREATH Study

- We have found that self-reported alcohol consumption decreases after ART initiation, however we suspect a high degree of under-report. We use the biomarker PEth to obtain a biological measure of alcohol use in a one-year prospective study. Looking at baseline and 3 months…

Comparison of two means: *paired* t-test

- We think participants may be decreasing their alcohol use over time. The null hypothesis is that they are drinking the same amount.

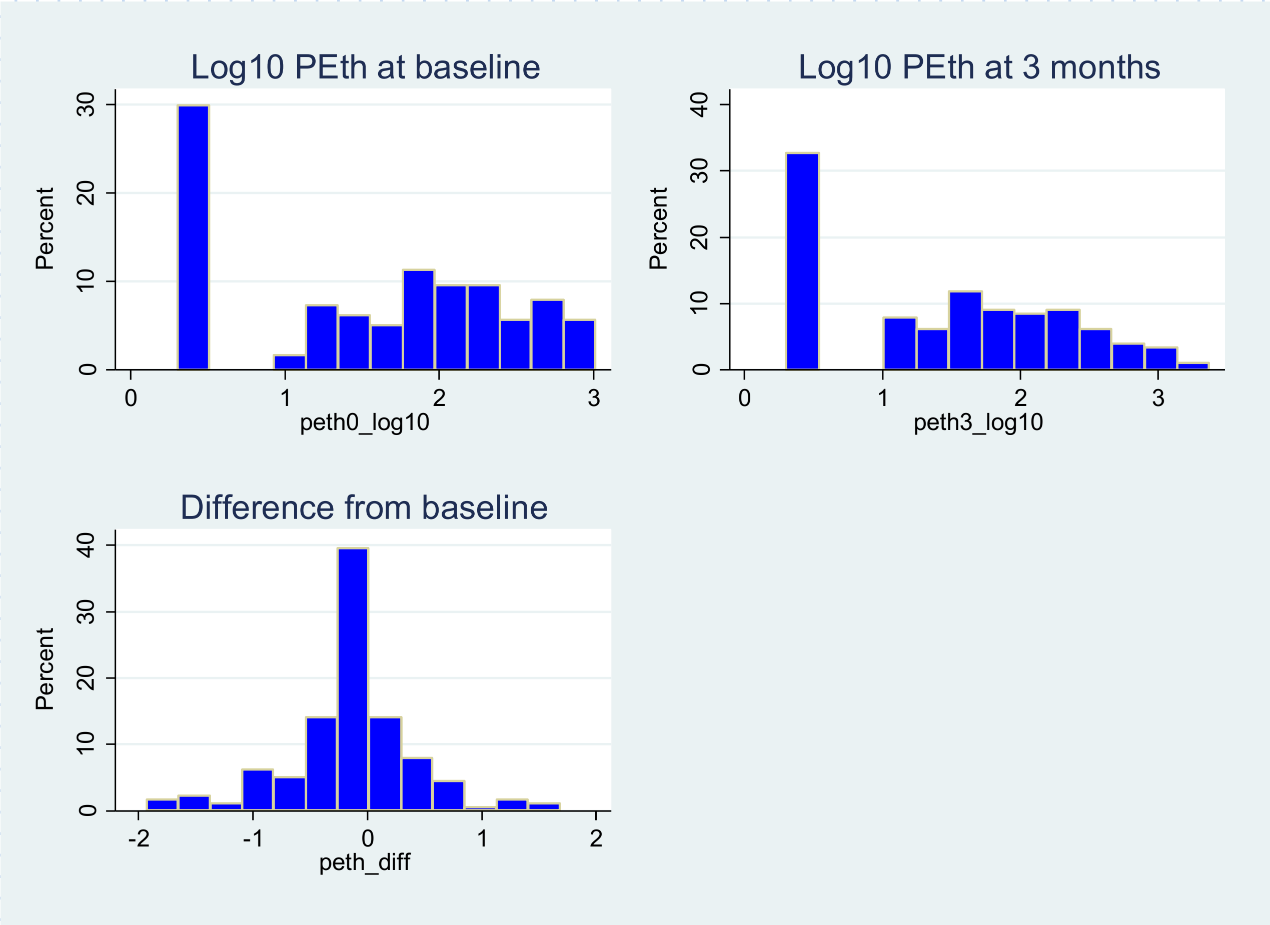
  H_0: μ_2-μ_1 =0 $\rightarrow$ μ_2=μ_1

  H_A: μ_1-μ_2 ≠0 $\rightarrow$ μ_2 ≠μ_1

- Significance level=0.05

Comparison of two means: *paired* t-test

Graphs of the data



The mean of the paired difference is a new random variable

Comparison of two means: *paired* t-test

```
. summm peth_diff

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
   peth_diff |       177    -.101553    .5782935  -1.929419   1.685742


*** calculate the t statistic
 di -.101553/.5782935*sqrt(177)
-2.3363133


*** calculate the p-value
. di 2*ttail(176,2.336133)
.02061082
```

$$t_{stat} = \frac{\bar{d}}{s_d / \sqrt{n}} \text{ where}$$

$$s_d = \sqrt{\frac{\sum_{i=1}^{n}(d_i - \bar{d})^2}{n-1}}$$

→ So we reject the null

Comparison of two means: *paired* t-test
Using the ttest command

```
. . ttest peth_diff==0

One-sample t test
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
peth_d~f |     177    -.101553    .0434672    .5782935    -.187337    -.015769
------------------------------------------------------------------------------
    mean = mean(peth_diff)                                        t =  -2.3363
Ho: mean = 0                                     degrees of freedom =      176

    Ha: mean < 0                 Ha: mean != 0                 Ha: mean > 0
 Pr(T < t) = 0.0103         Pr(|T| > |t|) = 0.0206          Pr(T > t) = 0.9897
```

Note that mean>0 here is *mean difference*

Comparison of two means: *paired* t-test

Another way without calculating the difference

The command is

```
ttest var1==var2
```

```
. . ttest peth3_log10==peth0_log10

Paired t test
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
peth3~10 |     177    1.426734    .0686534    .9133744    1.291244    1.562224
peth0~10 |     177    1.528287    .0690468    .9186077    1.392021    1.664553
---------+--------------------------------------------------------------------
    diff |     177   -.101553    .0434672    .5782935   -.187337    -.015769
------------------------------------------------------------------------------
     mean(diff) = mean(peth3_log10 - peth0_log10)                 t =  -2.3363
 Ho: mean(diff) = 0                              degrees of freedom =      176

 Ha: mean(diff) < 0           Ha: mean(diff) != 0           Ha: mean(diff) > 0
 Pr(T < t) = 0.0103         Pr(|T| > |t|) = 0.0206          Pr(T > t) = 0.9897

    .
```

Comparison of two *independent* means

- The goal is to compare means from two independent samples
- Two different populations
  - E.g. vaccine versus placebo group
  - E.g. women with adequate versus inadequate micronutrient levels
- Even though the null and alternative hypotheses are the same as for the paired t-test, the test is different, it is wrong to use a paired t-test with independent samples and vice versa

Comparison of two *independent* means

- So we can calculate a t or a z-score for the difference in the means and compare it to the standard normal distribution. The test statistics are:

$$z_{stat} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\sigma^2 \left(1/n_1 + 1/n_2\right)}}$$

$$t_{stat} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{s_p^2 \left(1/n_1 + 1/n_2\right)}} \text{ where}$$

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

Comparison of two *independent* means

- Study of non-pneumatic anti-shock garment (Miller et al)
- Two groups – pre-intervention received usual treatment, intervention group received NASG
- Comparison of hemorrhaging in the two groups
- Null hypothesis: The hemorrhaging is the same in the two groups $H_0: \mu_1 = \mu_2$

  $H_A: \mu_1 \neq \mu_2$
- The data: External blood loss after entry:
  - Pre-intervention group (n=83) mean blood loss =340.4 SD=248.2
  - Intervention group (n=83) mean blood loss =73.5 SD=93.9

Comparison of two means, example

*** ttesti n1 mean1 sd1 n2 mean2 sd2**

```
ttesti 83 340.4 248.2 83 73.5 93.9

Two-sample t test with equal variances
------------------------------------------------------------------------------
         |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       x |      83       340.4    27.24349       248.2     286.204     394.596
       y |      83        73.5    10.30686        93.9    52.99636    94.00364
---------+--------------------------------------------------------------------
combined |     166      206.95    17.85377    230.0297    171.6987    242.2013
---------+--------------------------------------------------------------------
    diff |               266.9    29.12798                209.3858    324.4142
------------------------------------------------------------------------------
    diff = mean(x) - mean(y)                                      t =   9.1630
Ho: diff = 0                                     degrees of freedom =      164

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

Comparison of two *independent* means

- You can calculate a 95% confidence interval for the difference between the 2 means

$$\text{Lower bound}: (\bar{x}_1 - \bar{x}_2) - t_{df,.025}\sqrt{s_p^2\left[\frac{1}{n_1} + \frac{1}{n_2}\right]}$$

$$\text{Upper bound}: (\bar{x}_1 - \bar{x}_2) + t_{df,.025}\sqrt{s_p^2\left[\frac{1}{n_1} + \frac{1}{n_2}\right]}$$

- If the confidence interval for the difference does not include 0, then you can reject the null hypothesis of no difference

Comparison of two *independent* means

- This t-test assumes equal variances in the two underlying populations
- If we do not assume equal variances we use a slightly different test statistic
    - Variances not assumed to be equal, so you do not use a pooled estimate
    - There is another formula for degrees of freedom
- Often the two different t-tests yield the same answer, but you should not assume equivalence unless you have a good reason for it
    - If the sample sizes are equal, you will get the same test statistic, just the df changes

Comparison of two means, example

```
ttesti 83 340.4 248.2 83 73.5 93.9, unequal

Two-sample t test with unequal variances
------------------------------------------------------------------------------
         |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       x |      83       340.4    27.24349       248.2     286.204     394.596
       y |      83        73.5    10.30686        93.9    52.99636    94.00364
---------+--------------------------------------------------------------------
combined |     166      206.95    17.85377    230.0297    171.6987    242.2013
---------+--------------------------------------------------------------------
    diff |               266.9    29.12798                209.1446    324.6554
------------------------------------------------------------------------------
    diff = mean(x) - mean(y)                                      t =   9.1630
Ho: diff = 0                     Satterthwaite's degrees of freedom =  105.002

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

Comparison of two *independent* means

- When you have the data in Stata, with the different groups in different columns, use

```
ttest  var1==var2, unpaired
```

or

```
ttest  var1==var2, unpaired unequal
```

- More commonly you will have the data all in one variable, and the grouping in another variable. Then use

```
ttest  var, by(groupvar)
```

or

```
ttest var, by(groupvar) unequal
```

Comparison of two *independent* means

Alcohol study example

Testing whether log PEth at 6 months differs by study arm

Null hypothesis: Log PEth for cohort arm = Log PEth for comparison arm

```
. ttest peth_log10, by(studyarm)

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
Cohort a |     181    1.406739    .0709392    .9543891     1.26676    1.546718
Min asse |     106    1.490502    .0897145    .9236671    1.312615    1.668389
---------+--------------------------------------------------------------------
combined |     287    1.437676    .0556285     .942407    1.328183    1.547169
---------+--------------------------------------------------------------------
    diff |           -.0837633    .1153577               -.3108244    .1432978
------------------------------------------------------------------------------
    diff = mean(Cohort a) - mean(Min asse)                        t =  -0.7261
Ho: diff = 0                                     degrees of freedom =      285

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.2342         Pr(|T| > |t|) = 0.4684          Pr(T > t) = 0.7658
```

Comparison of two *independent* means

Alcohol study example

Testing whether log PEth at 6 months differs by study arm

Null hypothesis: Log PEth for cohort arm = Log PEth for comparison arm

```
.
 ttest peth_log10, by(studyarm) unequal

Two-sample t test with unequal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
Cohort a |     181    1.406739    .0709392    .9543891     1.26676    1.546718
Min asse |     106    1.490502    .0897145    .9236671    1.312615    1.668389
---------+--------------------------------------------------------------------
combined |     287    1.437676    .0556285     .942407    1.328183    1.547169
---------+--------------------------------------------------------------------
    diff |           -.0837633    .1143724               -.3091369    .1416103
------------------------------------------------------------------------------
    diff = mean(Cohort a) - mean(Min asse)                        t =  -0.7324
Ho: diff = 0                     Satterthwaite's degrees of freedom =  225.846

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.2324         Pr(|T| > |t|) = 0.4647          Pr(T > t) = 0.7676
```

Comparison of two proportions

- Null hypothesis about two proportions, p_1 and p_2,

  H_0: $p_1 = p_2$

  H_A: $p_1 \neq p_2$

- If n_1 and n_2 are sufficiently large, the difference between the two proportions follows a normal distribution.
- So we can use the z statistic to find the probability of observing a difference as large as we do, under the null hypothesis of no difference

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\hat{p}(1-\hat{p})[1/n_1 + 1/n_2]}}$$

$$\text{where } \hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$$

Comparison of two proportions

- Example: Proportion with PEth>=50 ng/ml

```
. tab peth_50

   peth_50 |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |        155       54.01       54.01
          1 |        132       45.99      100.00
------------+-----------------------------------
      Total |        287      100.00


. bysort studyarm: tab peth_50
-------------------------------------------------------------------------------------->
   studyarm = Cohort assessed

    peth_50 |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |         97       53.59       53.59
          1 |         84       46.41      100.00
------------+-----------------------------------
      Total |        181      100.00
-----------------------------------------------------------------------------------
-> studyarm = Min assessed

    peth_50 |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |         58       54.72       54.72
          1 |         48       45.28      100.00
------------+-----------------------------------
      Total |        106      100.00


.
```

Comparison of two proportions

prtesti n1 p1 n2 p2

```
.prtesti 106 .453 181 .464

Two-sample test of proportions                     x: Number of obs =      106
                                                   y: Number of obs =      181
------------------------------------------------------------------------------
    Variable |       Mean   Std. Err.       z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |       .453   .0483493                      .3582372    .5477628
           y |       .464   .0370683                      .3913476    .5366524
-------------+----------------------------------------------------------------
        diff |      -.011   .0609238                     -.1304084    .1084084
             |  under Ho:   .0609565    -0.18   0.857
------------------------------------------------------------------------------
        diff = prop(x) - prop(y)                                  z =  -0.1805
    Ho: diff = 0

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(Z < z) = 0.4284         Pr(|Z| < |z|) = 0.8568          Pr(Z > z) = 0.5716
```

Comparison of two proportions

```
. prtest peth_50, by(month)

Two-sample test of proportions                     6: Number of obs =       181
                                                  88: Number of obs =       106
------------------------------------------------------------------------------
    Variable |       Mean   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           6 |   .4640884   .0370687                       .391435    .5367418
          88 |   .4528302   .0483477                      .3580704    .5475899
-------------+----------------------------------------------------------------
        diff |   .0112582   .0609228                     -.1081483    .1306647
             |  under Ho:   .0609564     0.18   0.853
------------------------------------------------------------------------------
        diff = prop(6) - prop(88)                                 z =   0.1847
    Ho: diff = 0

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(Z < z) = 0.5733         Pr(|Z| < |z|) = 0.8535          Pr(Z > z) = 0.4267

.

.
```

Types of error

- Type I error – significance level of the test
  α=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 α is set to 0.05.

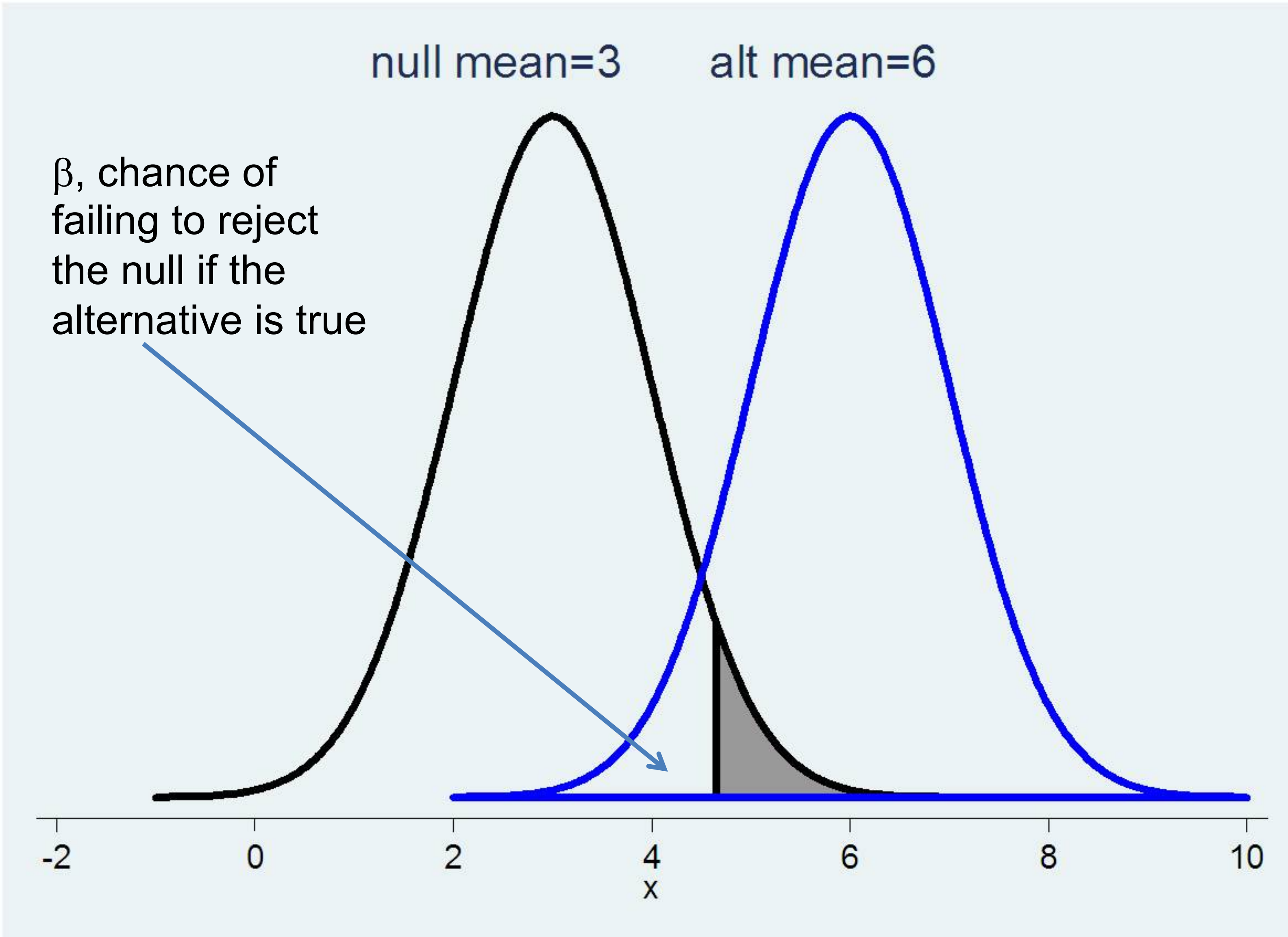
Types of error

- Type II error – β = 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*

	True state	
Decision	**H_0 is true**	**H_0 is false**
Do not reject H_0	Correct	Type II error=β
Reject H_0	Type I error=α	Correct

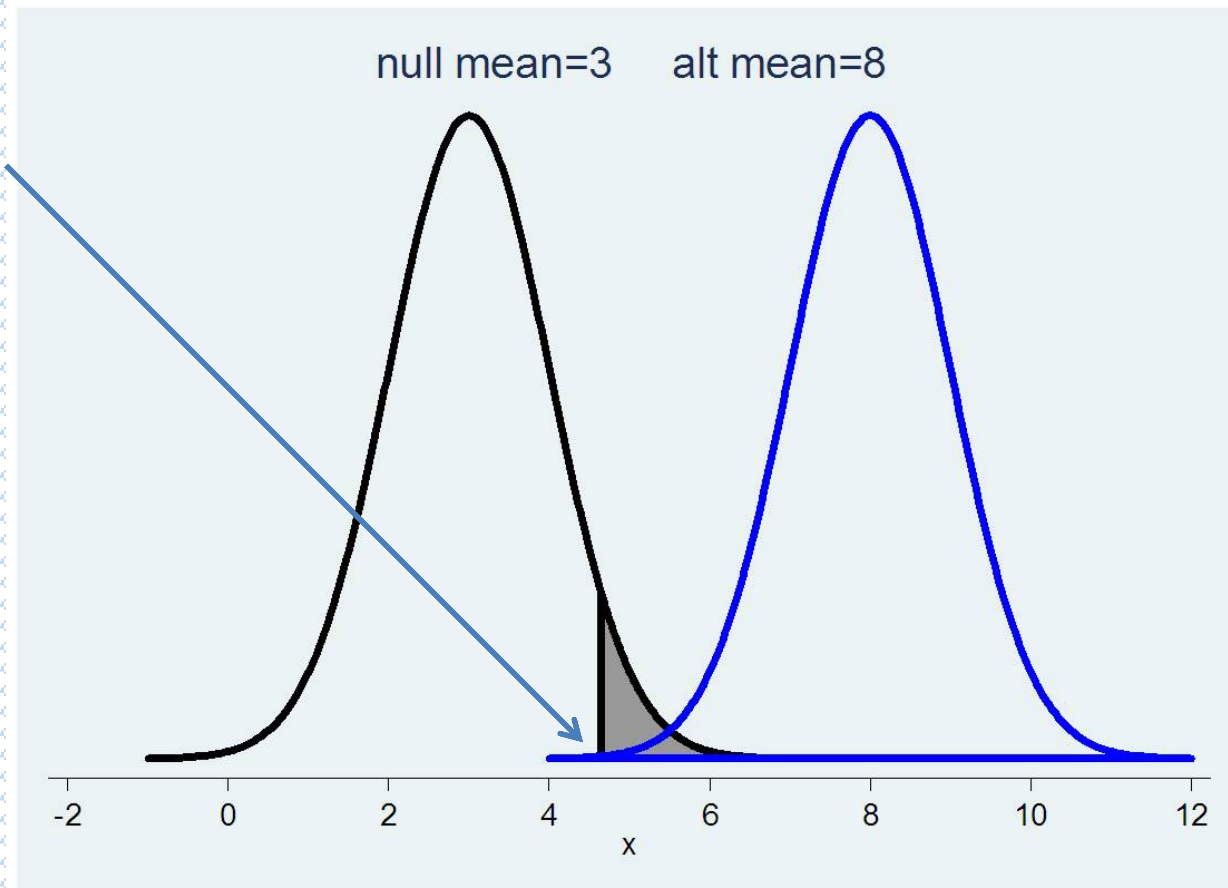
Chance of a type II error



Chance of a type II error

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

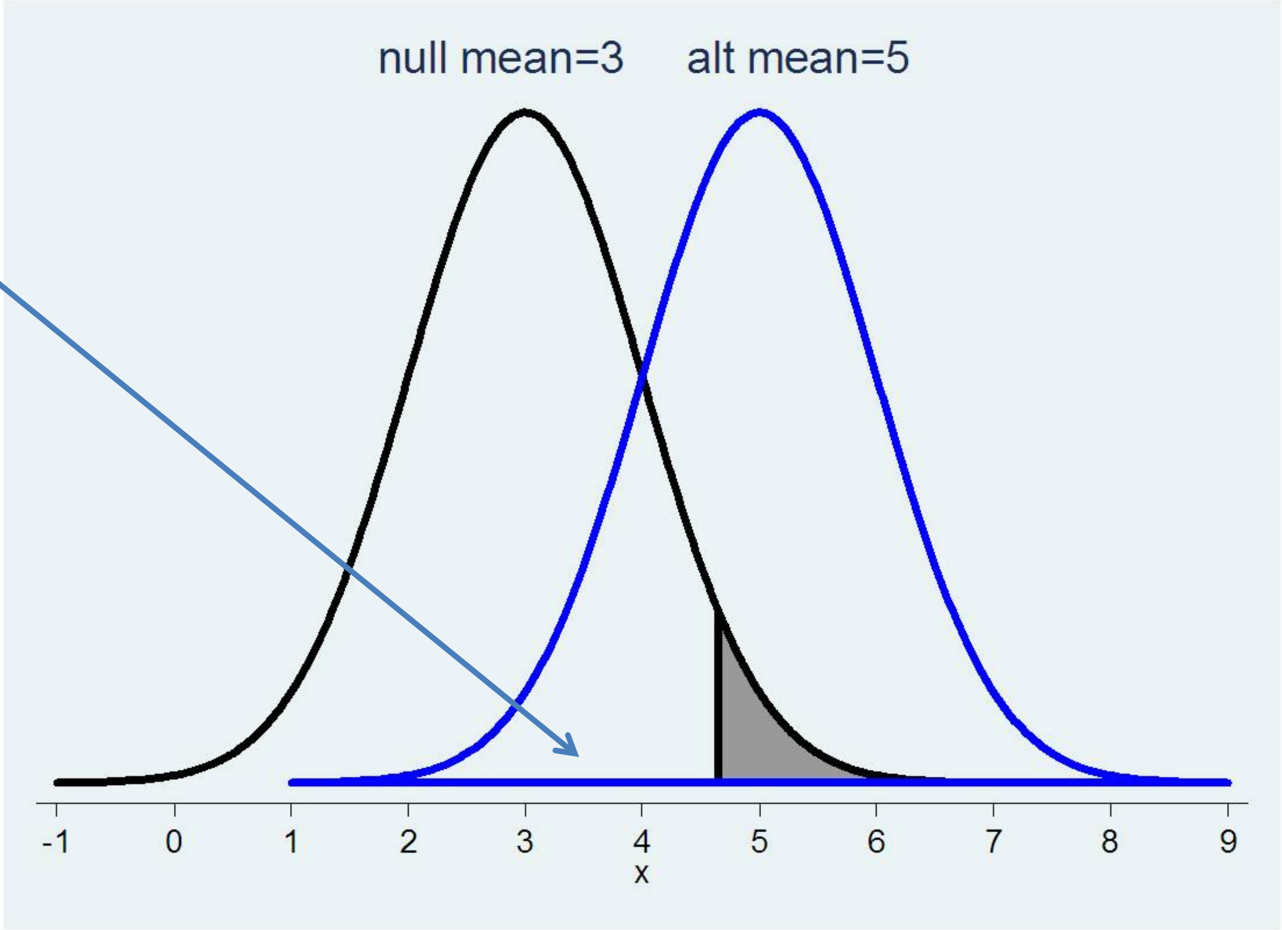
β, chance of failing to reject the null if the alternative is true



Chance of a type II error

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

β, chance of failing to reject the null if the alternative is true



Finding β, P(Type II error)

- Example: Mean age of walking
  - H_0: μ< 11.4 months (μ_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

Finding β, P(Type II error)

- If the true mean μ is really 14 meaning the null is false, what is the probability that the null will not be rejected? 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

Finding β, P(Type II error)

- 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-β is called the <u>power</u> of a statistical test

- The power of a test is the probability that you will reject the null hypothesis when you should

- 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 hypothesis value.
  - When the alternate mean was 14, power =1-.012=.988
  - When the alternate 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 α, 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 α and β levels.