Basic Statistics: MBI-540

Sally W. Thurston, PhD

Associate Professor
Department of Biostatistics and Computational Biology
University of Rochester

January 2011

Why statistics?

- Allows us to say something ("make inference") about a population based on data from a sample.
- Example: measure effect of drug on weight loss in 20 BALB/c mice
  - Sample: the 20 mice
  - Population: all BALB/c mice (that could be subject to the same condition)
  - Interested in weight loss in BALB/c mice, not just in the 20 mice.
  - Statistics allows us to estimate weight loss in the population using data from a sample.
- Many journal articles ask for statistics

Some comments

- Design of experiment is key
  - No statistical method can fix bad data!
  - Essential to think about design before data is collected
- Be clear about research question(s) before collecting data.
  - Includes: what data to collect to answer question(s).
- Know your data
  - How the measurements relate to research question
  - Plots to visualize your data: helpful!
- Most statistical models have assumptions
  - If violated, statistical “answer” not correct.

How can a statistician help?

- Design of experiment
  - To help, statistician must understand aims, background
- Analysis
  - Statistician can figure out appropriate model, check assumptions, fit model to data
- Research?
  - Talk with statistician early: help with design to prevent bias
  - If small project, consultation possible
  - If detailed project, statistician should be collaborator, not consultant

Outline

Dr. Tuba's question

Dr. Tuba is interested in the effect of drugs on weight loss in mice after infection.

- Drugs are given right after infection
- 6 mice were given the standard drug
- 6 mice were given the new drug
- Dr. Tuba reports $p < 0.05$ for a t-test of the difference in mean weight loss across the two groups after day 5.

Dr. Tuba's question

Dr. Tuba is interested in the effect of drugs on weight loss in mice after infection.

- Drugs are given right after infection
- 6 mice were given the standard drug
- 6 mice were given the new drug
- Dr. Tuba reports $p < 0.05$ for a t-test of the difference in mean weight loss across the two groups after day 5.

Should Dr. Tuba publish?

Dr. Tuba's question

Dr. Tuba is interested in the effect of drugs on weight loss in mice after infection.

- Drugs are given right after infection
- 6 mice were given the standard drug
- 6 mice were given the new drug
- Dr. Tuba reports $p < 0.05$ for a t-test of the difference in mean weight loss across the two groups after day 5.

Should Dr. Tuba publish?

- Your thoughts?
- What additional information might you need before deciding?

Some things to consider

- How did Dr. Tuba choose which mice received each drug?
- Are the mice independent?
- Plot the data - why?
- Did he choose to present results only from the time period that showed the most significant effect?

How did Dr. Tuba choose the mice?

Scenario A

- The 12 mice were running around in a pen.
- Dr. Tuba captured the mice one at a time.
- The first 6 mice captured were given the standard drug.

How did Dr. Tuba choose the mice?

Scenario A

- The 12 mice were running around in a pen.
- Dr. Tuba captured the mice one at a time.
- The first 6 mice captured were given the standard drug.

Problems?

How did Dr. Tuba choose the mice?

Scenario A

- The 12 mice were running around in a pen.
- Dr. Tuba captured the mice one at a time.
- The first 6 mice captured were given the standard drug.

Problems?

- Ease of capture might be associated with propensity to lose weight.
  - Fastest mice might be the most healthy
  - Fastest mice might be the lightest
- So - mouse selection could be the cause of an apparent drug effect.

How did Dr. Tuba choose the mice?

Scenario A

- The 12 mice were running around in a pen.
- Dr. Tuba captured the mice one at a time.
- The first 6 mice captured were given the standard drug.

Problems?

- Ease of capture might be associated with propensity to lose weight.
  - Fastest mice might be the most healthy
  - Fastest mice might be the lightest
- So - mouse selection could be the cause of an apparent drug effect.

Important to randomize treatment assignment.

How did Dr. Tuba choose the mice?

Scenario B

- The mice came from 2 dams, treatment was randomized.
- Pups from the first dam were given the standard drug.
- Pups from the second dam were given the new drug.

How did Dr. Tuba choose the mice?

Scenario B

- The mice came from 2 dams, treatment was randomized.
- Pups from the first dam were given the standard drug.
- Pups from the second dam were given the new drug.

Problems?

How did Dr. Tuba choose the mice?

Scenario B

- The mice came from 2 dams, treatment was randomized.
- Pups from the first dam were given the standard drug.
- Pups from the second dam were given the new drug.

Problems?

- The effect of drug is totally confounded with litter
- Translation: If we see a difference in outcome between the two groups, we can't tell if it was
  - caused by the drug - **OR**
  - due to a litter effect
- Mice in one litter might have a different propensity to lose weight than mice in another litter.

How did Dr. Tuba choose the mice?

Scenario C

- The mice came from 6 dams.
- The first pup in each dam was given the standard drug.
- The second pup in each dam was given the new drug.

How did Dr. Tuba choose the mice?

Scenario C

- The mice came from 6 dams.
- The first pup in each dam was given the standard drug.
- The second pup in each dam was given the new drug.

Problems?

How did Dr. Tuba choose the mice?

Scenario C

- The mice came from 6 dams.
- The first pup in each dam was given the standard drug.
- The second pup in each dam was given the new drug.

Problems?

- Birth order may be associated with the propensity to lose weight.
- Birth order is confounded with drug.

How did Dr. Tuba choose the mice?

Scenario C

- The mice came from 6 dams.
- The first pup in each dam was given the standard drug.
- The second pup in each dam was given the new drug.

Problems?

- Birth order may be associated with the propensity to lose weight.
- Birth order is confounded with drug.

Better idea: within a litter, randomize first/second born pups to drug.

How did Dr. Tuba choose the mice?

Scenario D

- Mice were all first-born pups from different dams.
- Mice were randomized to treatment
- All mice which received the standard drug were housed in one cage.
- All mice which received the new drug were housed in one (different) cage.

How did Dr. Tuba choose the mice?

Scenario D

- Mice were all first-born pups from different dams.
- Mice were randomized to treatment
- All mice which received the standard drug were housed in one cage.
- All mice which received the new drug were housed in one (different) cage.

Problems?

How did Dr. Tuba choose the mice?

Scenario D

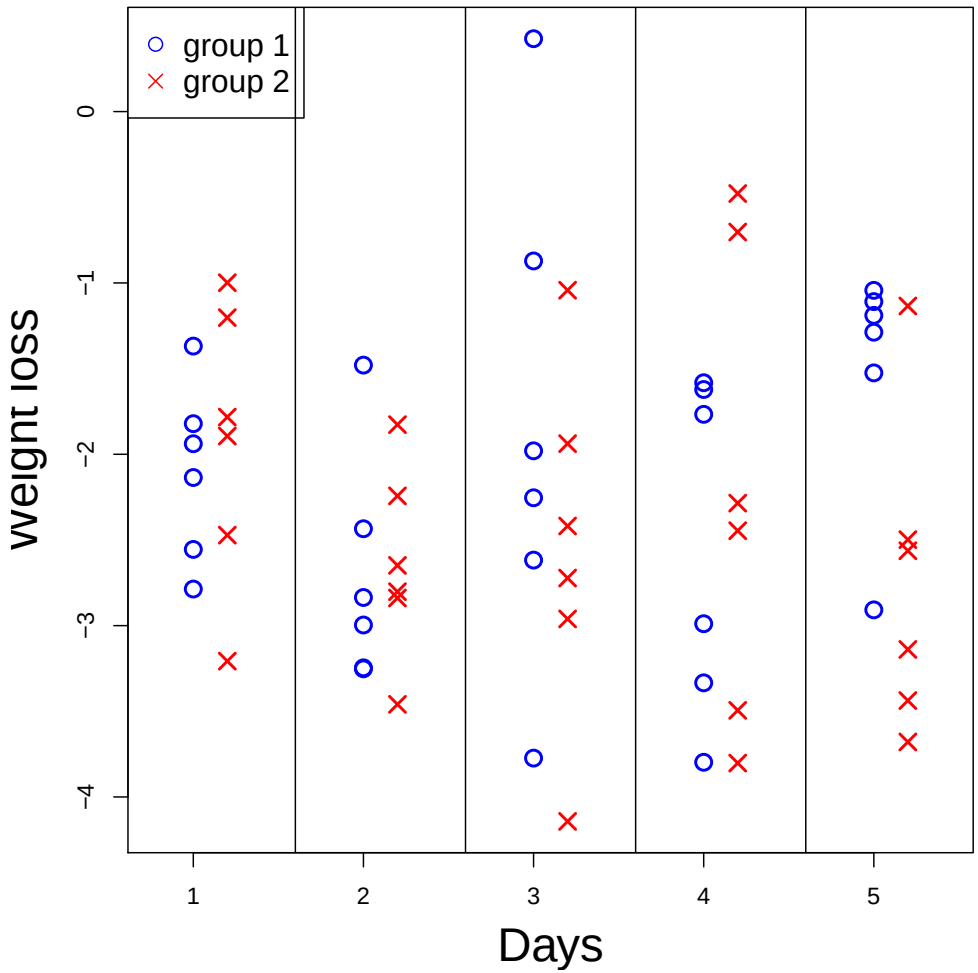
- Mice were all first-born pups from different dams.
- Mice were randomized to treatment
- All mice which received the standard drug were housed in one cage.
- All mice which received the new drug were housed in one (different) cage.

Problems?

- There might be a cage effect - viral infection, hotter location, less water, $\cdots$.

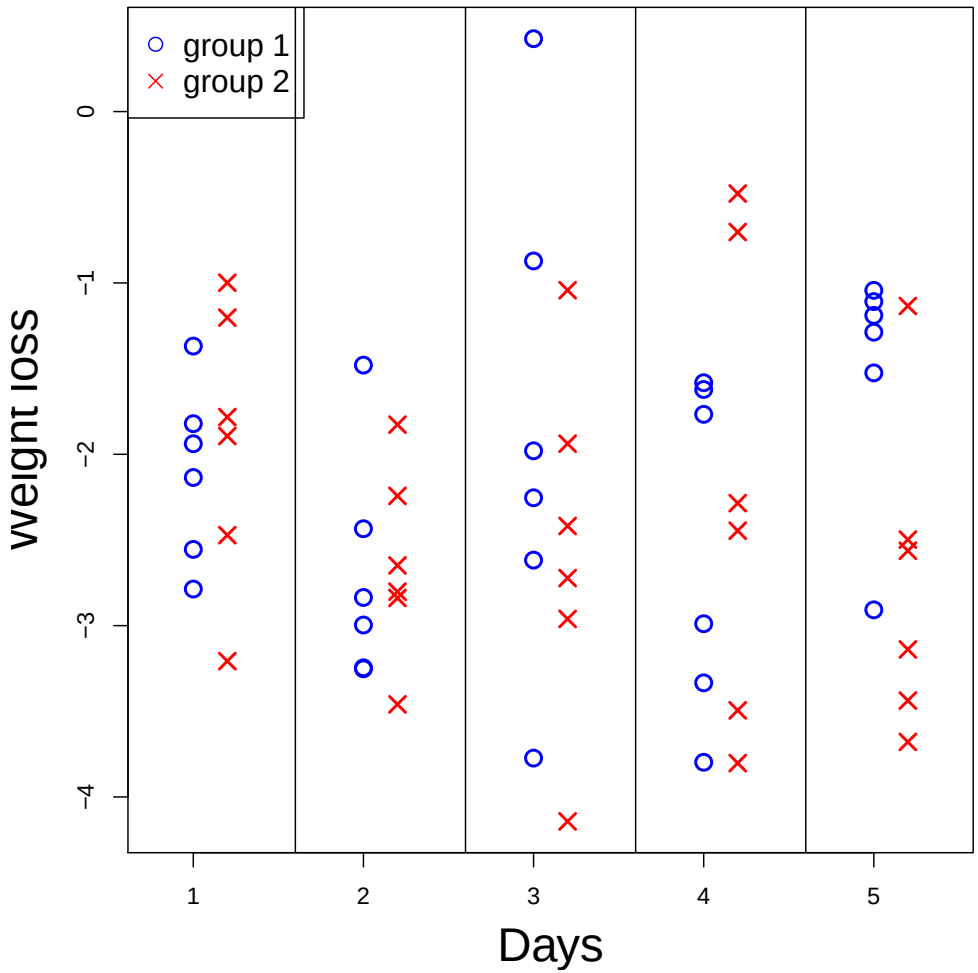
Better idea: divide animals from a drug group into $>$ 1 cage - or if necessary, house all in same cage.

Was day 5 one of many days he considered?



Comments?

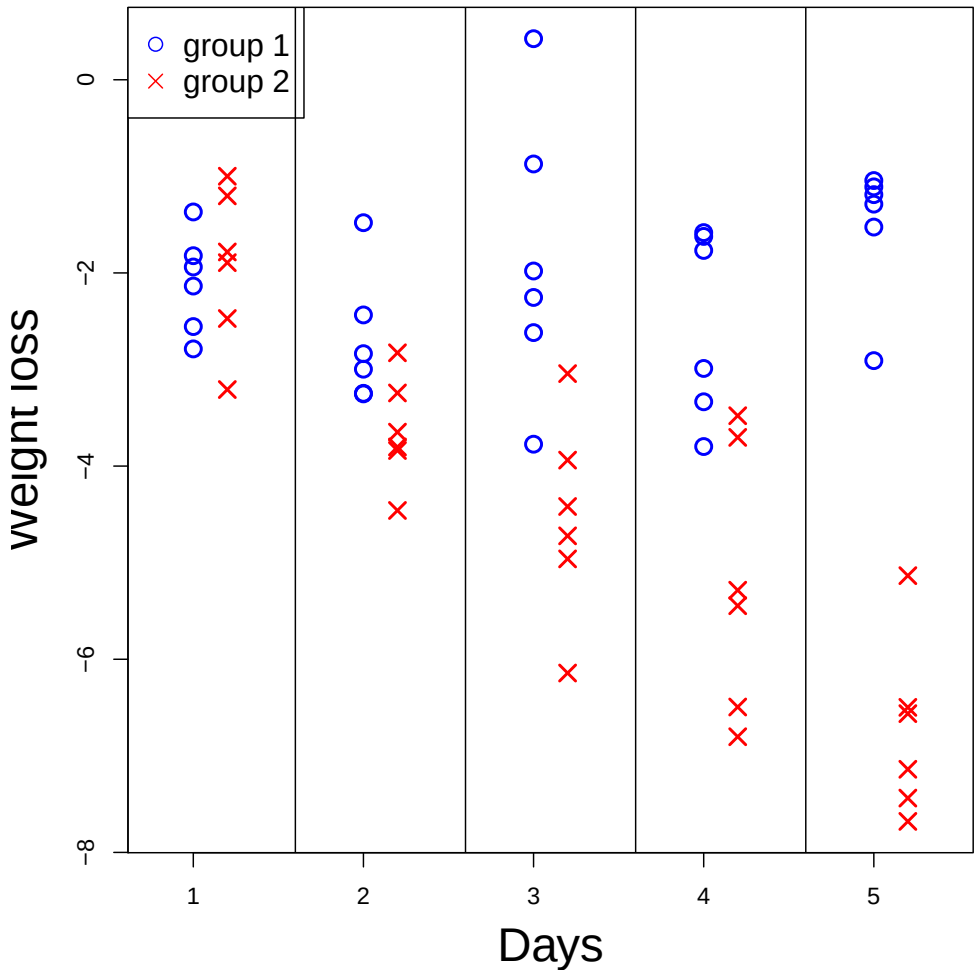
Was day 5 one of many days he considered?



Comments?

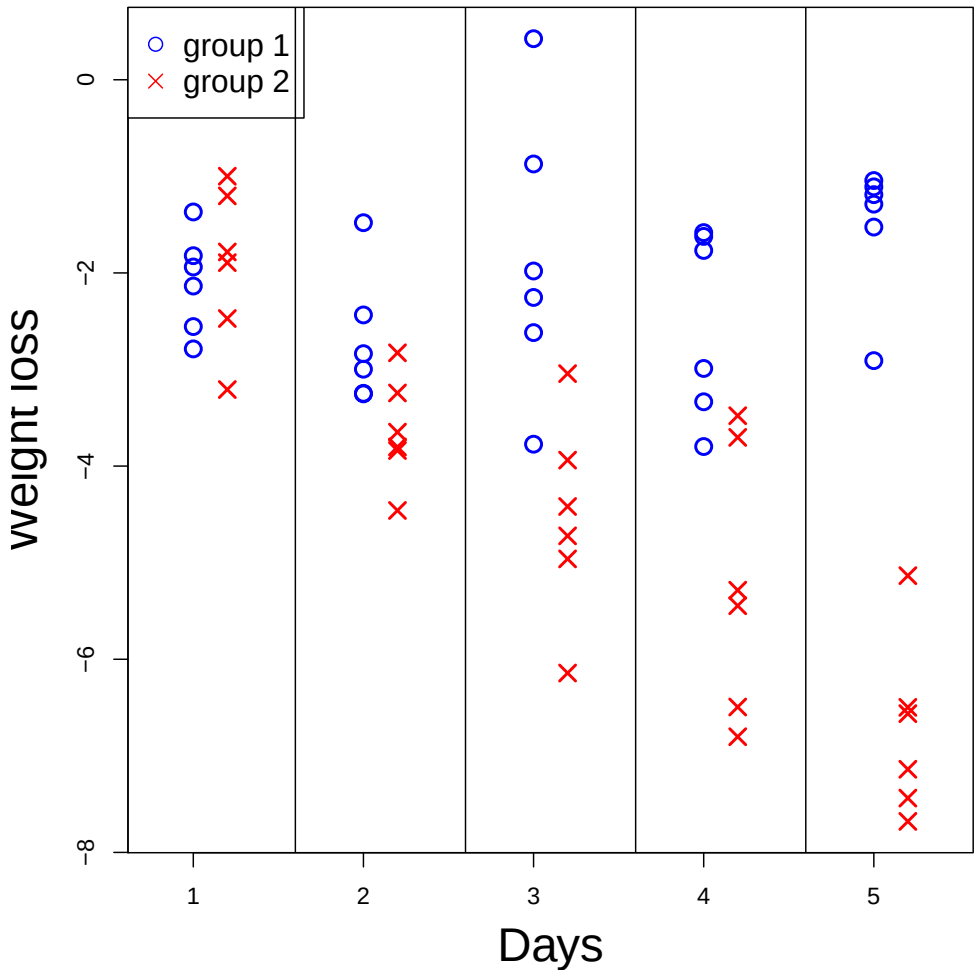
- Simulated data: the mean is the same for both groups, at all times
- Looking for the biggest difference = multiple comparisons

Was day 5 one of many days he considered?



Comments?

Was day 5 one of many days he considered?



Comments?

- Simulated data: the mean declines with time for group 2, not group 1
- If expect this, could choose day 5 *a priori*
- Could use other statistical methods to compare slope over time in the 2 groups.

Are the mice independent?

- Mice from the same litter are more like each other.
- Mice housed in the same cage might respond more similarly
- We are not interested in the effect of litter (or cage)

Are the mice independent?

- Mice from the same litter are more like each other.
- Mice housed in the same cage might respond more similarly
- We are not interested in the effect of litter (or cage)

If appropriate experimental design, can control for correlated observations statistically.
Need to record and save data on the litter (or cage)

Other scenarios (1)

An investigator is comparing response to different doses of 3 drugs on 96-well plates.

- Each drug is tested at 5 doses.
- Plate 1 has multiple wells of drug A at each dose level.
- Plate 2 has multiple wells of drug B at each dose level.
- Plate 3 has multiple wells of drug C at each dose level.

Other scenarios (1)

An investigator is comparing response to different doses of 3 drugs on 96-well plates.

- Each drug is tested at 5 doses.
- Plate 1 has multiple wells of drug A at each dose level.
- Plate 2 has multiple wells of drug B at each dose level.
- Plate 3 has multiple wells of drug C at each dose level.

Is this a good experimental design? Why / why not?

Other scenarios (1)

An investigator is comparing response to different doses of 3 drugs on 96-well plates.

- Each drug is tested at 5 doses.
- Plate 1 has multiple wells of drug A at each dose level.
- Plate 2 has multiple wells of drug B at each dose level.
- Plate 3 has multiple wells of drug C at each dose level.

Is this a good experimental design? Why / why not?

If one drug appears to be better, it might be due to a difference in the plates.

Other scenarios (2)

Dr. Busy scores mouse activity in a 1-minute period on a scale of 1-10.

- All mice are measured 6 hours after treatment
- 10 mice receive a standard treatment, 10 a new treatment.
- Treatment assignment is randomized
- The new treatment is expected to be better.
- The investigator knows the treatment assignment of each mouse.

Other scenarios (2)

Dr. Busy scores mouse activity in a 1-minute period on a scale of 1-10.

- All mice are measured 6 hours after treatment
- 10 mice receive a standard treatment, 10 a new treatment.
- Treatment assignment is randomized
- The new treatment is expected to be better.
- The investigator knows the treatment assignment of each mouse.

Is this a good experimental design? Why / why not?

Other scenarios (2)

Dr. Busy scores mouse activity in a 1-minute period on a scale of 1-10.

- All mice are measured 6 hours after treatment
- 10 mice receive a standard treatment, 10 a new treatment.
- Treatment assignment is randomized
- The new treatment is expected to be better.
- The investigator knows the treatment assignment of each mouse.

Is this a good experimental design? Why / why not?

Expectation that new treatment is better could unconsciously bias the investigator's measurements

Comments on experimental designs

- In the scenarios just described, there is a possibility that the intended effect cannot be distinguished from something else that is not of interest.
- Estimate of intended effect may be biased.

Comments on experimental designs

- In the scenarios just described, there is a possibility that the intended effect cannot be distinguished from something else that is not of interest.
- Estimate of intended effect may be biased.
- Unless the unintended effect can be estimated (impossible?)

No statistical analysis will fix this problem.

Comments on experimental designs

- In the scenarios just described, there is a possibility that the intended effect cannot be distinguished from something else that is not of interest.
- Estimate of intended effect may be biased.
- Unless the unintended effect can be estimated (impossible?)

  No statistical analysis will fix this problem.
- Often, the problem can be avoided by good experimental design.

Summary: experimental design

- Data should be a random sample from population
- If treatment is assigned, assignment should be randomly determined.
  - Can use table of random numbers
- Blinding
  - Person measuring endpoints should not know treatment assignment
  - Person receiving treatment should not know his/her treatment
  - Particularly an issue if measurements can be subjective

Outline

Dr. Forte's question

- A well-known journal article reports that after infection, BALB/c mice given 'superdrug' lost 4 ounces on average after one week.
- Dr. Forte wants to compare this to 'superdrug' effects on infected C57BL/6 mice.
- Dr. Forte infects 4 C57BL/6 mice, treats with 'superdrug', and measures weight loss after one week.
  - His measurements of weight loss: 5, 8, 6, 3 ounces.

The mean

- Data (weight loss) represented by $x_1, x_2, \cdots, x_n$
  - n is sample size (number observations)
  - Dr. Forte's data: $x_1 = 5, x_2 = 8, x_3 = 6, x_4 = 3$.
- Population mean (μ): the average x_i over all observations in population
  - Cannot measure data from all observations in population.
  - μ is unknown (a population parameter)
- Estimate μ by $\hat{\mu}$. Here $\hat{\mu} = \bar{x}$ = sample mean.

The mean

- Data (weight loss) represented by $x_1, x_2, \cdots, x_n$
  - n is sample size (number observations)
  - Dr. Forte's data: $x_1 = 5, x_2 = 8, x_3 = 6, x_4 = 3$.
- Population mean (μ): the average x_i over all observations in population
  - Cannot measure data from all observations in population.
  - μ is unknown (a population parameter)
- Estimate μ by $\hat{\mu}$. Here $\hat{\mu} = \bar{x}$ = sample mean.

For Dr. Forte's experiment

- What is n?
- What is the sample?
- What is the population?

The mean (cont.)

- Dr. Forte's data: $3, 5, 6, 8$.
- Sample mean, $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$

What is Dr. Forte's sample mean?

The mean (cont.)

- Dr. Forte's data: $3, 5, 6, 8$.
- Sample mean, $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$

What is Dr. Forte's sample mean?

$\bar{x} = \frac{1}{4}(3 + 5 + 6 + 8) = 5.5$

Variability

Consider measurements from 2 researchers:

- Dr. Forte's measurements: $3, 5, 6, 8$.
- Dr. Pianissimo's measurements: $5.1, 5.3, 5.7, 5.9$
- Both have $\bar{x} = 5.5$.

Whose measurements give you more confidence about your knowledge of the population mean?

Variability

Consider measurements from 2 researchers:

- Dr. Forte's measurements: $3, 5, 6, 8$.
- Dr. Pianissimo's measurements: $5.1, 5.3, 5.7, 5.9$
- Both have $\bar{x} = 5.5$.

Whose measurements give you more confidence about your knowledge of the population mean?

Dr. Pianissimo: data closer together, plausible values for μ in smaller interval.

Variance (σ^2 = population variance)

- Population variance: $\sigma^2 = \mathrm{E}[(x_i - \mu)^2]$
  - E means "expectation": average over all observations in population.
  - Translation: σ^2 is expected squared difference between individual values and population mean
  - σ = population SD (of x_i)
- Sample variance, $\hat{\sigma}^2$
  - Estimate σ^2 by $\hat{\sigma}^2$.
  - If μ known: $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$

Variance (σ^2 = population variance)

- Population variance: $\sigma^2 = \mathrm{E}[(x_i - \mu)^2]$
  - E means "expectation": average over all observations in population.
  - Translation: σ^2 is expected squared difference between individual values and population mean
  - σ = population SD (of x_i)
- Sample variance, $\hat{\sigma}^2$
  - Estimate σ^2 by $\hat{\sigma}^2$.
  - If μ known: $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$
  - If μ unknown: $\hat{\sigma}^2 = s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$

Dr. Forte's data: $3, 5, 6, 8$, $\bar{x} = 5.5$. What is s^2?

Variance (σ^2 = population variance)

- Population variance: $\sigma^2 = \mathrm{E}[(x_i - \mu)^2]$
  - E means "expectation": average over all observations in population.
  - Translation: σ^2 is expected squared difference between individual values and population mean
  - σ = population SD (of x_i)
- Sample variance, $\hat{\sigma}^2$
  - Estimate σ^2 by $\hat{\sigma}^2$.
  - If μ known: $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$
  - If μ unknown: $\hat{\sigma}^2 = s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$

Dr. Forte's data: $3, 5, 6, 8$, $\bar{x} = 5.5$. What is s^2?

$s^2 = \frac{(3-5.5)^2+(5-5.5)^2+(6-5.5)^2+(8-5.5)^2}{3} = 4.333$

Variance (σ^2 = population variance)

- Population variance: $\sigma^2 = \mathrm{E}[(x_i - \mu)^2]$
  - E means "expectation": average over all observations in population.
  - Translation: σ^2 is expected squared difference between individual values and population mean
  - σ = population SD (of x_i)
- Sample variance, $\hat{\sigma}^2$
  - Estimate σ^2 by $\hat{\sigma}^2$.
  - If μ known: $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$
  - If μ unknown: $\hat{\sigma}^2 = s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$

Dr. Forte's data: $3, 5, 6, 8$, $\bar{x} = 5.5$. What is s^2?
$s^2 = \frac{(3-5.5)^2+(5-5.5)^2+(6-5.5)^2+(8-5.5)^2}{3} = 4.333$

Dr. Pianissimo's data: $5.1, 5.3, 5.7, 5.9$, $s^2 = 0.133$.

SD, SE

- Standard deviation (of x), SD(x) is $s = \sqrt{s^2}$
  - Can also write s as s_x
- Standard error (of $\bar{x}$), SE($\bar{x}$) is $\sqrt{\frac{s^2}{n}} = \frac{s}{\sqrt{n}}$
- In other contexts, can speak of standard error of other quantities, e.g. SE($\hat{\beta}$)

SD, SE

- Standard deviation (of x), SD(x) is $s = \sqrt{s^2}$
  - Can also write s as s_x
- Standard error (of $\bar{x}$), SE($\bar{x}$) is $\sqrt{\frac{s^2}{n}} = \frac{s}{\sqrt{n}}$
- In other contexts, can speak of standard error of other quantities, e.g. SE($\hat{\beta}$)

Dr. Forte's data: $3, 5, 6, 8$,
$\bar{x} = 5.5, s^2 = 4.333$.

What are SD(x) and SE($\bar{x}$)?

SD, SE

- Standard deviation (of x), SD(x) is $s = \sqrt{s^2}$
  - Can also write s as s_x
- Standard error (of $\bar{x}$), SE($\bar{x}$) is $\sqrt{\frac{s^2}{n}} = \frac{s}{\sqrt{n}}$
- In other contexts, can speak of standard error of other quantities, e.g. SE($\hat{\beta}$)

Dr. Forte's data: $3, 5, 6, 8$,
$\bar{x} = 5.5, s^2 = 4.333$.

What are SD(x) and SE($\bar{x}$)?

- SD(x) $= \sqrt{4.333} = 2.08$
- SE($\bar{x}$) $= \frac{2.08}{\sqrt{4}} = 1.04$

Dr. Pianissimo's data: $5.1, 5.3, 5.7, 5.9$
$\bar{x} = 5.5, s^2 = 0.133$.

What are SD(x) and SE($\bar{x}$)?

SD, SE

- Standard deviation (of x), SD(x) is $s = \sqrt{s^2}$
  - Can also write s as s_x
- Standard error (of $\bar{x}$), SE($\bar{x}$) is $\sqrt{\frac{s^2}{n}} = \frac{s}{\sqrt{n}}$
- In other contexts, can speak of standard error of other quantities, e.g. SE($\hat{\beta}$)

Dr. Forte's data: $3, 5, 6, 8$,
$\bar{x} = 5.5, s^2 = 4.333$.

What are SD(x) and SE($\bar{x}$)?

- SD(x) $= \sqrt{4.333} = 2.08$
- SE($\bar{x}$) $= \frac{2.08}{\sqrt{4}} = 1.04$

Dr. Pianissimo's data: $5.1, 5.3, 5.7, 5.9$
$\bar{x} = 5.5, s^2 = 0.133$.

What are SD(x) and SE($\bar{x}$)?

- SD(x) $= \sqrt{0.133} = 0.365$
- SE($\bar{x}$) $= \frac{0.365}{\sqrt{4}} = 0.183$

SD, SE (cont.)

- Recall SD(x) is $s_x = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$, SE($\bar{x}$) $= \frac{s_x}{\sqrt{n}}$

If Dr. Forte had collected data from 20 mice instead of 4, how would SD(x) and SE($\bar{x}$) be affected?

SD, SE (cont.)

- Recall SD(x) is $s_x = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$, SE($\bar{x}$) $= \frac{s_x}{\sqrt{n}}$

If Dr. Forte had collected data from 20 mice instead of 4, how would SD(x) and SE($\bar{x}$) be affected?

- SD(x) would be about the same (estimated more precisely)
- SE($\bar{x}$) would decrease (we have more confidence about $\bar{x}$).

SD, SE (cont.)

- Recall SD(x) is $s_x = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$, SE($\bar{x}$) $= \frac{s_x}{\sqrt{n}}$

If Dr. Forte had collected data from 20 mice instead of 4, how would SD(x) and SE($\bar{x}$) be affected?

- SD(x) would be about the same (estimated more precisely)
- SE($\bar{x}$) would decrease (we have more confidence about $\bar{x}$).

What do SD(x) and SE($\bar{x}$) mean?

- SD(x) is our estimate of σ, the SD of x in the population.
- SE($\bar{x}$) is a measure of our uncertainty about μ.
  - Can form confidence interval (CI) for μ: $\bar{x} \pm t_{crit}\text{SE}(\bar{x})$
  - t_{crit} discussed later (usually ≈ 2)
  - Approximately 95% of such CIs include μ.

Why divide by $n-1$ in s^2?

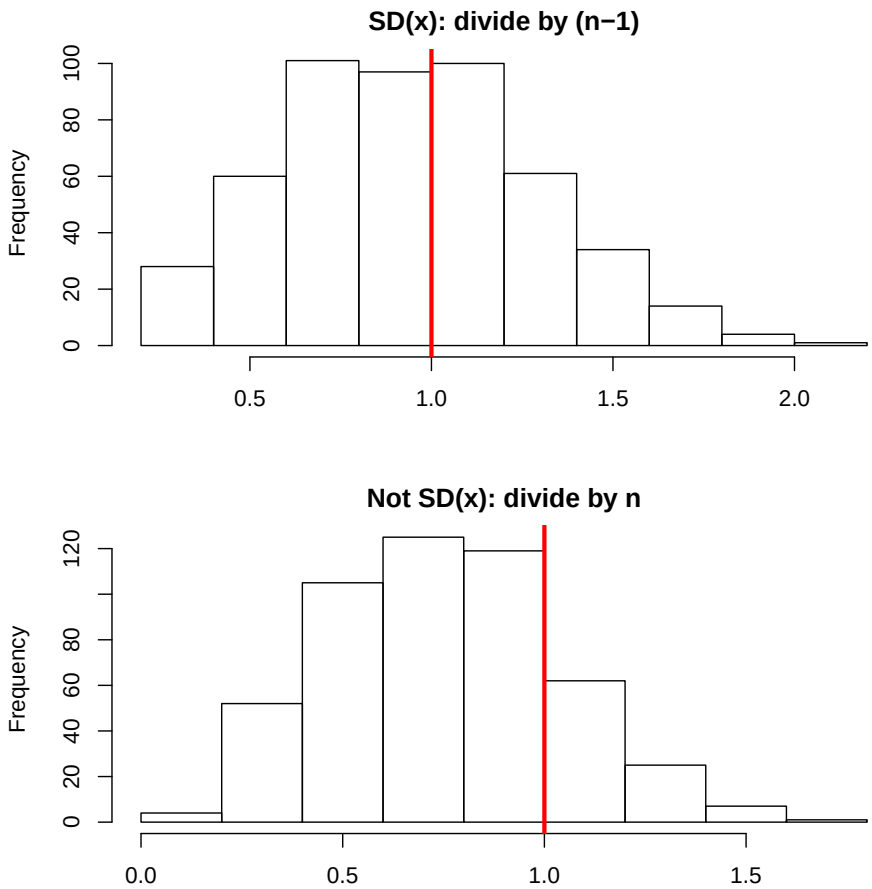
- If we knew μ, $\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{n}$
- We don't know μ, so $\hat{\sigma}^2 = s^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}$
- For a particular sample, $\bar{x}$ minimizes $\sum(x_i - \hat{\mu})^2$ for any $\hat{\mu}$
- Using $\bar{x}$ to estimate μ:
  - Numerate of s^2 is (on average) smaller than $\sum(x_i - \mu)^2$
  - Need to divide by smaller number to compensate; $n-1$ is the right number

Why divide by $n-1$ in s^2?

- Examine by simulation
- Take $n = 5$ observations from normal, $\mu = 0, \sigma^2 = 1$.
- Calculate $s^2 = \frac{\sum(x_i - \bar{x})^2}{4}$ and $\frac{\sum(x_i - \bar{x})^2}{5}$
- Repeat multiple times
- Note that true value of σ is 1.

Why divide by $n-1$ in s^2?

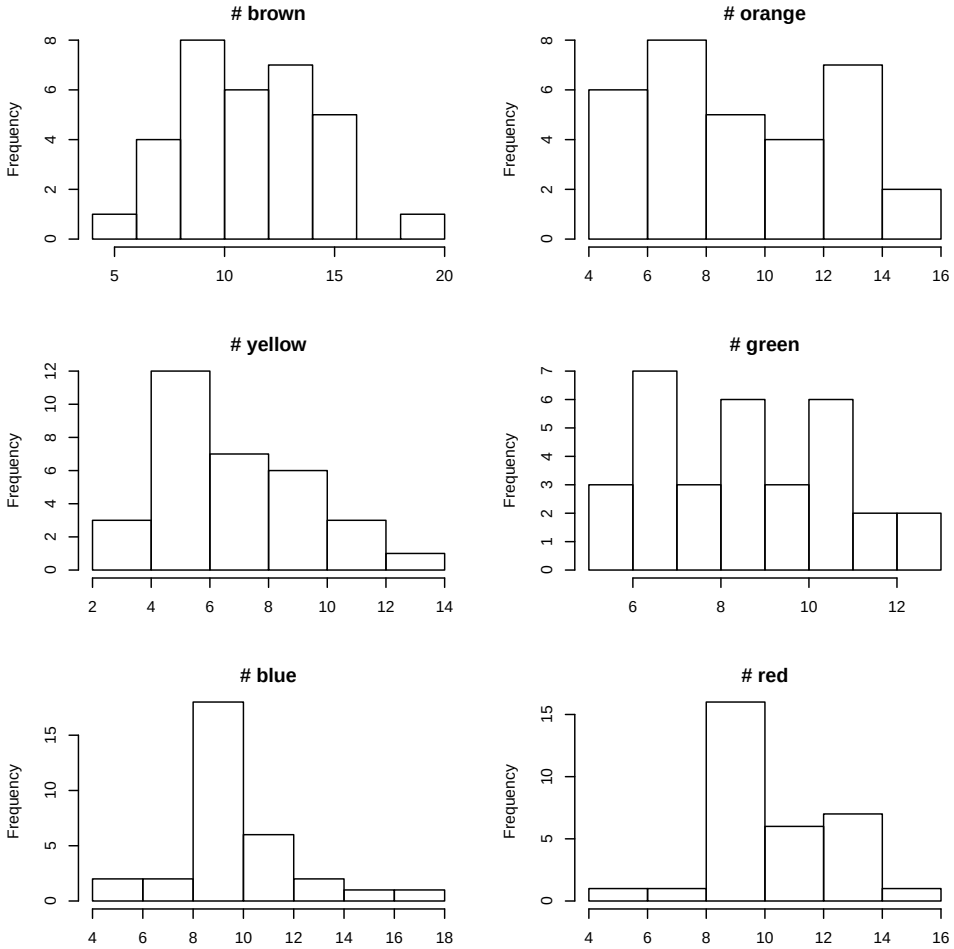
- Examine by simulation
- Take $n = 5$ observations from normal, $\mu = 0, \sigma^2 = 1$.
- Calculate $s^2 = \frac{\sum(x_i - \bar{x})^2}{4}$ and $\frac{\sum(x_i - \bar{x})^2}{5}$
- Repeat multiple times
- Note that true value of σ is 1.
- Dividing by $n-1$ does well
- Dividing by n underestimates σ



Outline

M&M data

We collected data on the number of M&M of each color in packages of mini M&M's



- 32 people total
- Histograms show number of each color

M&M data

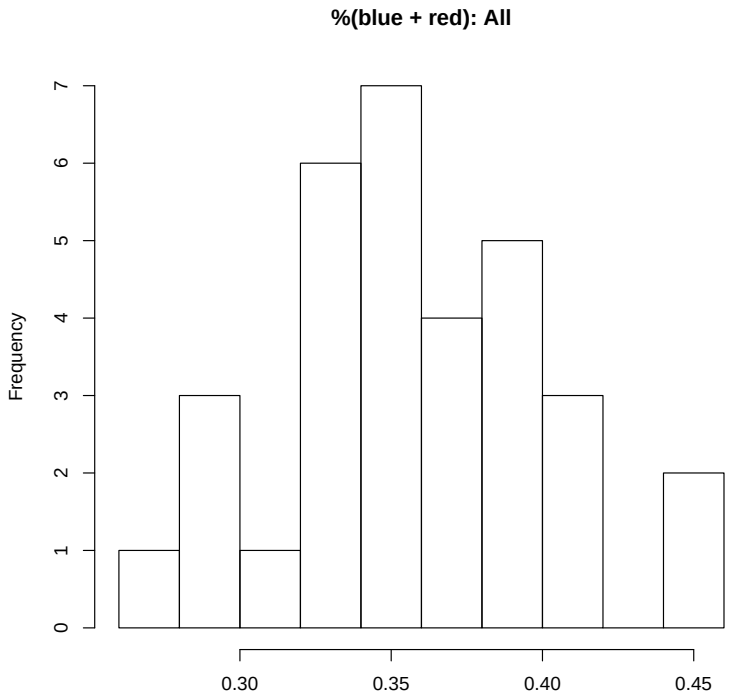
Aside: did everyone get same # M&M's? NO

Number MM:	50	51	55	56	57	58	59	60	61	62	66	73
People:	1	1	3	2	9	2	4	5	1	2	1	1

M&M data

Aside: did everyone get same # M&M's? NO

Number MM:	50	51	55	56	57	58	59	60	61	62	66	73
People:	1	1	3	2	9	2	4	5	1	2	1	1

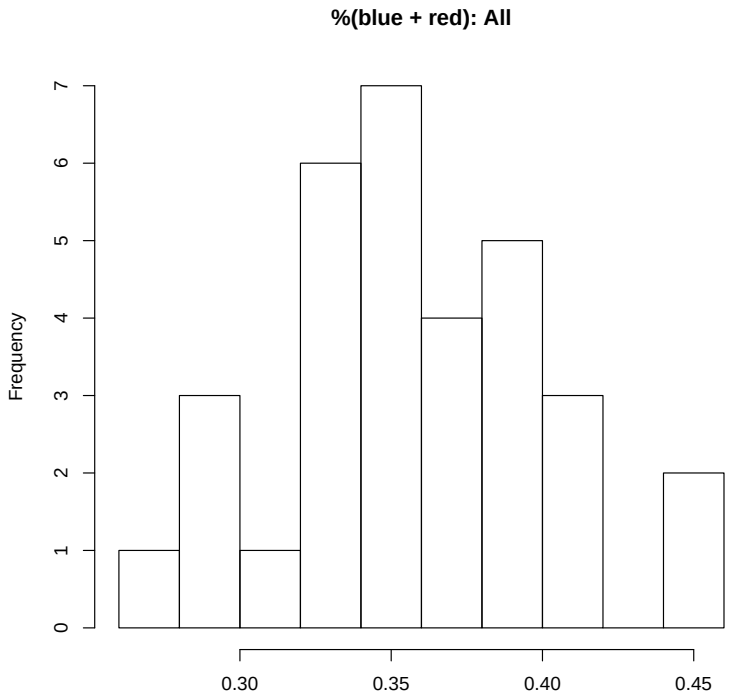


- Focus on % (red + blue)
- x_1 is % (red + blue) for first person, x_2 same for second person, etc.
- What is the sample?
- What is the population?

M&M data

Aside: did everyone get same # M&M's? NO

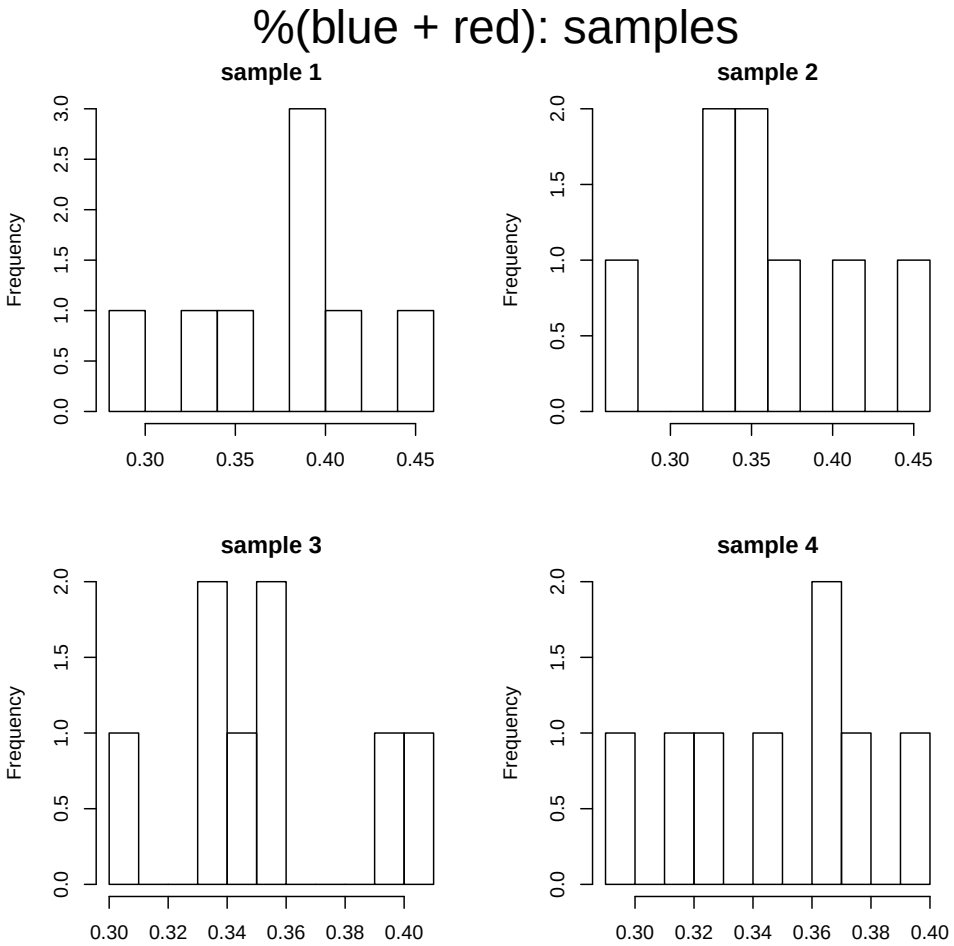
Number MM:	50	51	55	56	57	58	59	60	61	62	66	73
People:	1	1	3	2	9	2	4	5	1	2	1	1



- Focus on % (red + blue)
- x_1 is % (red + blue) for first person, x_2 same for second person, etc.
- What is the sample?
- What is the population?

We could approximate x_i with a normal distribution, mean μ, variance σ^2

M&M data



- Interest: % (red + blue)
- Divide data into 4 groups of 8 observations
- Calculate mean, SD, SE($\bar{x}$) in each group

M&M data

- Each group has $n = 8$ observations (32 observations total)
- Group 1: % (red + blue): Data $(x_1, \cdots, x_8)$ is

$$0.45, 0.38, 0.39, 0.38, 0.35, 0.33, 0.41, 0.29$$

- Calculate $\bar{x}$

M&M data

- Each group has $n = 8$ observations (32 observations total)
- Group 1: % (red + blue): Data $(x_1, \cdots, x_8)$ is

$$0.45, 0.38, 0.39, 0.38, 0.35, 0.33, 0.41, 0.29$$

- Calculate $\bar{x}$
  $\bar{x} = \frac{(0.45+0.38+0.39+0.38+0.35+0.33+0.41+0.29)}{8} = 0.374$
- Calculate s_x (standard deviation)

M&M data

- Each group has $n = 8$ observations (32 observations total)
- Group 1: % (red + blue): Data $(x_1, \cdots, x_8)$ is

$$0.45, 0.38, 0.39, 0.38, 0.35, 0.33, 0.41, 0.29$$

- Calculate $\bar{x}$
  $\bar{x} = \frac{(0.45+0.38+0.39+0.38+0.35+0.33+0.41+0.29)}{8} = 0.374$
- Calculate s_x (standard deviation)
  $s_x = \sqrt{\frac{(0.45-0.374)^2+(0.38-0.374)^2+\cdots+(0.29-0.374)^2}{7}} = 0.048$
- Calculate $\text{SE}(\bar{x}) = \frac{s_x}{\sqrt{n}} = \frac{0.048}{\sqrt{8}} = 0.017$

M&M data

Each group has $n = 8$ observations (32 observations total)

- Group 1: $\bar{x}$= 0.374, s_x= 0.0479 SE($\bar{x}$)= 0.0169
- Group 2: $\bar{x}$= 0.3564, s_x= 0.0574 SE($\bar{x}$)= 0.0203
- Group 3: $\bar{x}$= 0.3549, s_x= 0.0348 SE($\bar{x}$)= 0.0123
- Group 4: $\bar{x}$= 0.3465, s_x= 0.0342 SE($\bar{x}$)= 0.0121

M&M data

Each group has $n = 8$ observations (32 observations total)

- Group 1: $\bar{x}$= 0.374, s_x= 0.0479 SE($\bar{x}$)= 0.0169
- Group 2: $\bar{x}$= 0.3564, s_x= 0.0574 SE($\bar{x}$)= 0.0203
- Group 3: $\bar{x}$= 0.3549, s_x= 0.0348 SE($\bar{x}$)= 0.0123
- Group 4: $\bar{x}$= 0.3465, s_x= 0.0342 SE($\bar{x}$)= 0.0121

Everyone: $\bar{x}$= 0.358 s_x= 0.0436 SE($\bar{x}$)= 0.0077

M&M data

Each group has $n = 8$ observations (32 observations total)

- Group 1: $\bar{x}$= 0.374, s_x= 0.0479 SE($\bar{x}$)= 0.0169
- Group 2: $\bar{x}$= 0.3564, s_x= 0.0574 SE($\bar{x}$)= 0.0203
- Group 3: $\bar{x}$= 0.3549, s_x= 0.0348 SE($\bar{x}$)= 0.0123
- Group 4: $\bar{x}$= 0.3465, s_x= 0.0342 SE($\bar{x}$)= 0.0121

Everyone: $\bar{x}$= 0.358 s_x= 0.0436 SE($\bar{x}$)= 0.0077

We might want to know if % red + blue in the population = $\frac{1}{3}$
(later: t-test)

Outline

t-test

Recall Dr. Forte's question: **is weight loss in C57BL/6 mice (his data) different than in BALB/c mice (published data, weight loss = 4 ounces)?**

- H_0 is the null hypothesis ("no effect", "same as before")
- H_A is the alternative hypothesis ("some effect", "not H_0")

t-test

Recall Dr. Forte's question: **is weight loss in C57BL/6 mice (his data) different than in BALB/c mice (published data, weight loss = 4 ounces)?**

- H_0 is the null hypothesis ("no effect", "same as before")
- H_A is the alternative hypothesis ("some effect", "not H_0")
- Example: $H_0 : \mu = \mu_0$ versus $H_A : \mu \neq \mu_0$.
  - μ is mean weight loss for C57Bl/6 mice in population
  - μ is unknown, but we have an estimate of it ($\bar{x}$)
  - μ_0 is some specific number (here: 4).

t-test

Recall Dr. Forte's question: **is weight loss in C57BL/6 mice (his data) different than in BALB/c mice (published data, weight loss = 4 ounces)?**

- H_0 is the null hypothesis ("no effect", "same as before")
- H_A is the alternative hypothesis ("some effect", "not H_0")
- Example: $H_0 : \mu = \mu_0$ versus $H_A : \mu \neq \mu_0$.
  - μ is mean weight loss for C57Bl/6 mice in population
  - μ is unknown, but we have an estimate of it ($\bar{x}$)
  - μ_0 is some specific number (here: 4).
- How to evaluate H_0, H_A?
  - Compare $\bar{x}$ to μ_0
  - If $\bar{x}$ is far from μ_0, suggests H_0 not true.
  - Need SE($\bar{x}$) to evaluate how different $\bar{x}$ and μ_0 are.

t-test assumptions

- T-test is based on $t_{calc} = \frac{\bar{x} - \mu_0}{SE(\bar{x})}$
- Assumptions: $x_i \sim$ independent $N(\mu, \sigma^2)$
  - Interpretation: $\sim$ means "distributed as"
  - Says data are independent and have a normal distribution
  - μ = population mean (of the x_i's)
  - σ^2 = population variance (of the x_i's) (σ = SD)

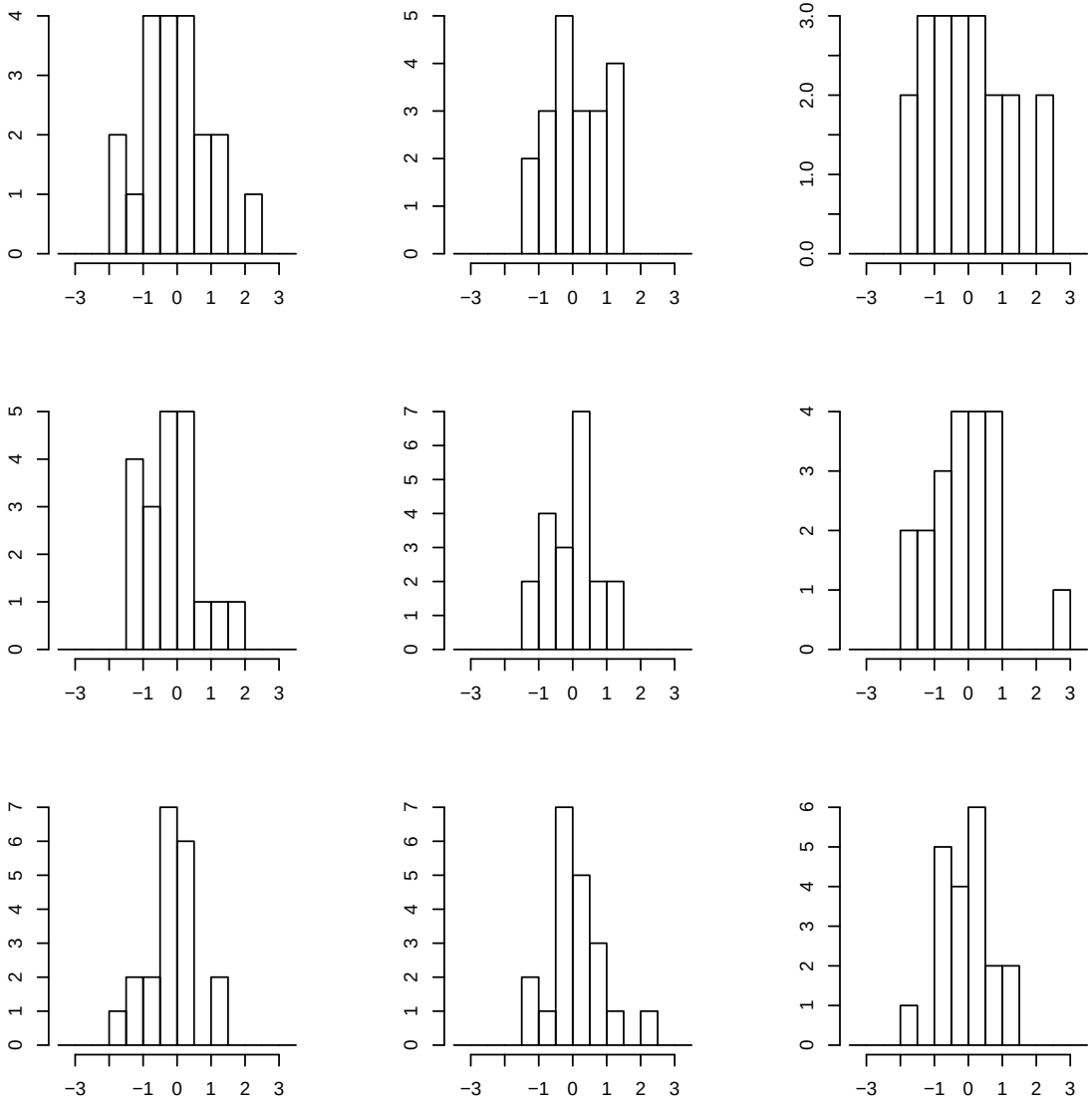
t-test assumptions

- T-test is based on $t_{calc} = \frac{\bar{x} - \mu_0}{SE(\bar{x})}$
- Assumptions: $x_i \sim$ independent $N(\mu, \sigma^2)$
  - Interpretation: $\sim$ means "distributed as"
  - Says data are independent and have a normal distribution
  - μ = population mean (of the x_i's)
  - σ^2 = population variance (of the x_i's) (σ = SD)

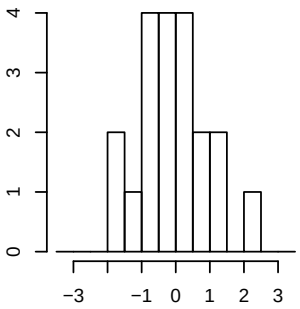
"How can I tell if my data have a normal distribution?"

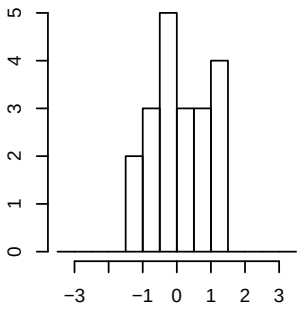
- Assumption is about distribution of data from population
- *t*-test robust to moderate departure from normality
- Important to plot data to check for gross violation from normality

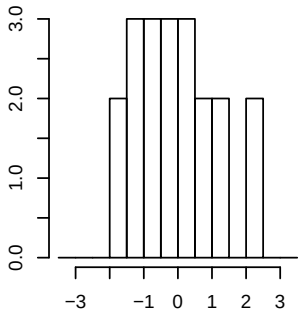
Samples of n=20 from a normal distribution

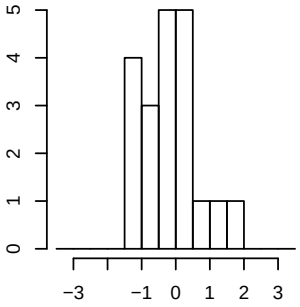


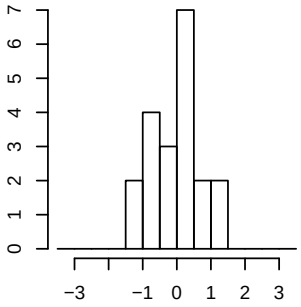
Samples of n=20 from a normal distribution

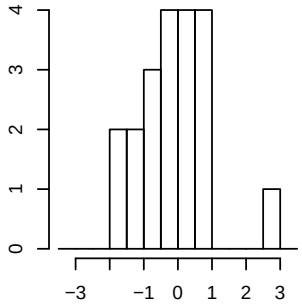


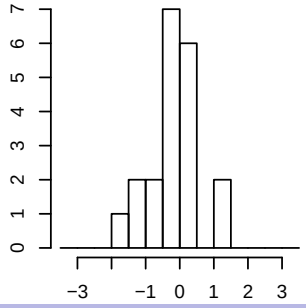


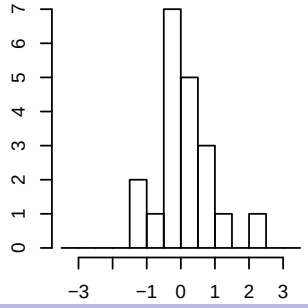


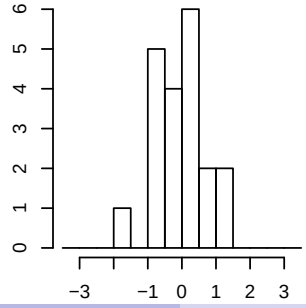




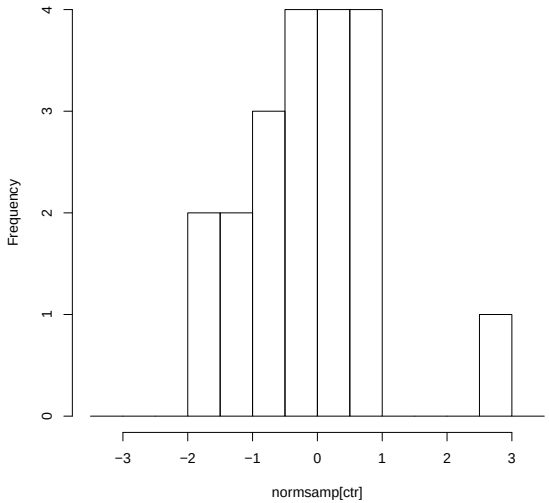




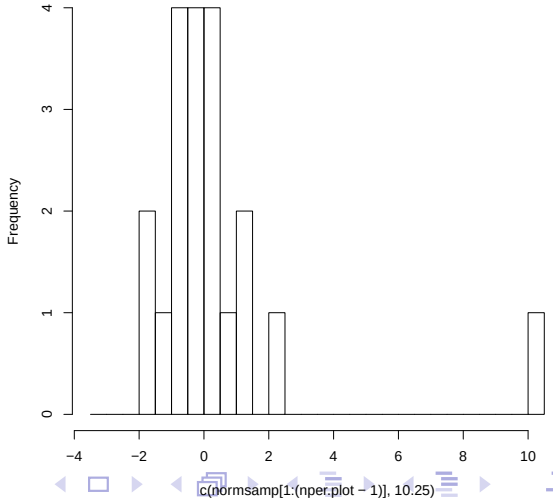




Normal



One mistake



Outliers

"My data has one very unusual value. Should I delete it?"

Outliers

"My data has one very unusual value. Should I delete it?"

- Step 1: check to see if the value is a mistake! If so, fix.
- Step 2: is there something unusual about that observation?
  - animal was sick
  - a different person took the measurement
  - experimental conditions different from others

Outliers

"My data has one very unusual value. Should I delete it?"

- Step 1: check to see if the value is a mistake! If so, fix.
- Step 2: is there something unusual about that observation?
  - animal was sick
  - a different person took the measurement
  - experimental conditions different from others
- If clear reason why unusual value differs from other observations
  - Value represents different condition: OK to delete (explain in paper)
- If cannot determine any reason why value is unusual
  - Deleting observation underestimates population variability: don't delete
  - Alternative: report results both with and without the value (explain in paper)

Parametric or non-parametric?

"When should I use a non-parametric version of a t-test?"

- Some statisticians (nearly) always use non-parametric tests (such as Mann-Whitney)
- Some statisticians (nearly) always use parametric tests (such as t-test)
- Knowledge of type of data can help determine if normality is reasonable
  - Sometimes data approximately normal after transformation
- If no severe departure from normality, either parametric or nonparametric tests probably fine.

t-test example

T-test is based on $t_{calc} = \frac{\bar{x} - \mu_0}{SE(\bar{x})}$

- Dr. Forte: $H_0 : \mu = 4, H_A : \mu \neq 4$.
- We calculated $\bar{x} = 5.5$, $SE(\bar{x}) = 1.04$.
- What is t_{calc}?

t-test example

T-test is based on $t_{calc} = \frac{\bar{x} - \mu_0}{SE(\bar{x})}$

- Dr. Forte: $H_0 : \mu = 4, H_A : \mu \neq 4$.
- We calculated $\bar{x} = 5.5$, $SE(\bar{x}) = 1.04$.
- What is t_{calc}?
  - Here $t_{calc} = (5.5 - 4)/1.04 = 1.44$

Under H_0, $t_{calc} \sim t_{n-1}$

- Here: $t_{calc} \sim t_3$
- If H_0 true, t_{calc} centered around 0

t-distribution

Recall $t_{calc} = \frac{\bar{x} - \mu_0}{SE(\bar{x})}$

- "How can t_{calc} have a distribution? It's just a number! "
- Your thoughts?

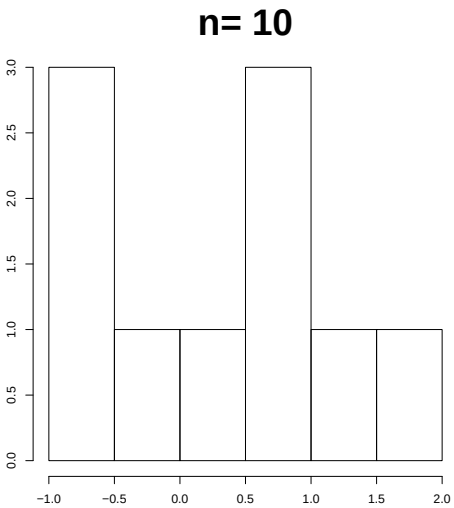
t-distribution

Recall $t_{calc} = \frac{\bar{x} - \mu_0}{SE(\bar{x})}$

- "How can t_{calc} have a distribution? It's just a number! "
- Your thoughts?

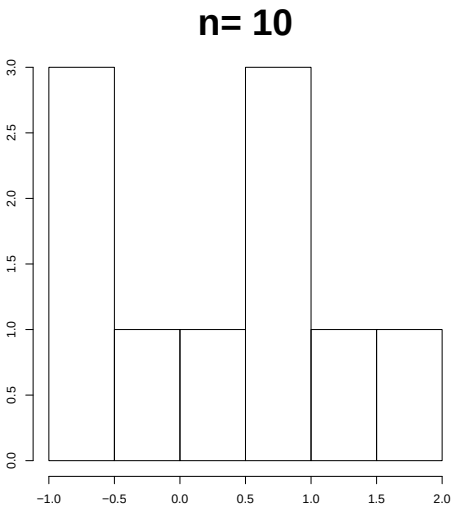
- Any single experiment gives a single t_{calc}.
- Could take another sample of 4 C57BL/6 mice, measure weight loss, get t_{calc} for that sample.
- This could theoretically be repeated multiple times with different mice
- Each experiment would give a t_{calc}
- A histogram of the t_{calc} values would look like the t_3 distribution (assuming enough experiments)

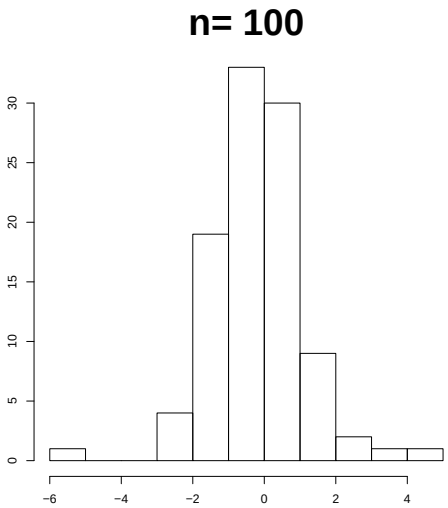
Samples from t-distributions with 3 df



- n=10: First few observations: -0.69, 1.73, 0.55, 0.96, -0.79, $\cdots$

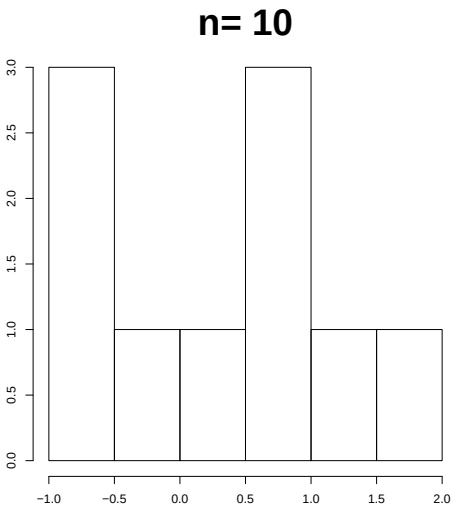
Samples from t-distributions with 3 df

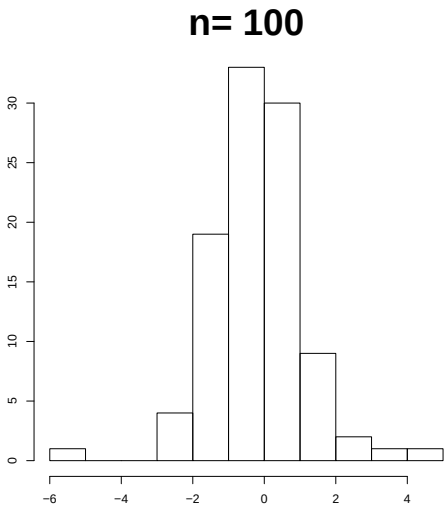


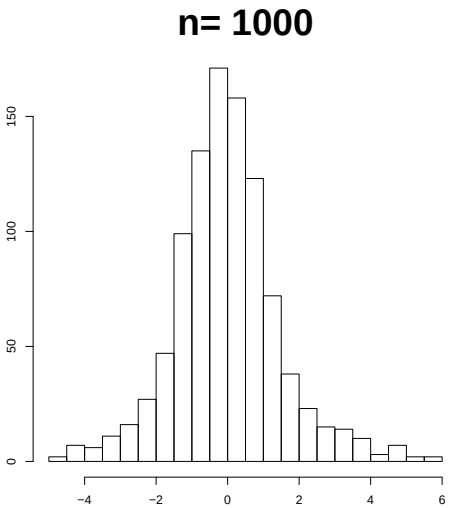


- n=10: First few observations: -0.69, 1.73, 0.55, 0.96, -0.79, $\cdots$
- n=100: Starting to look roughly "normal"

Samples from t-distributions with 3 df

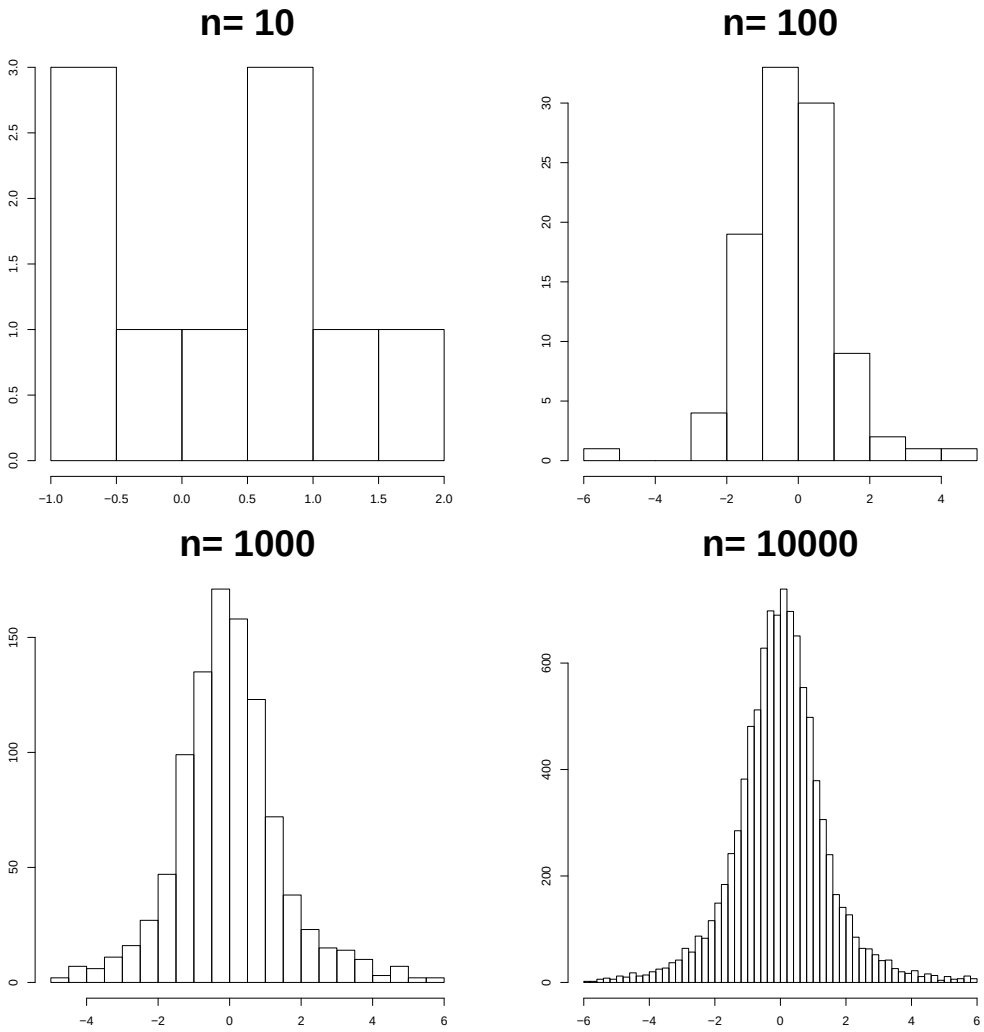






- n=10: First few observations: -0.69, 1.73, 0.55, 0.96, -0.79, $\cdots$
- n=100: Starting to look roughly “normal”
- n=1000: (12 obs outside plot. Range was: -12.8 to 8.3)

Samples from t-distributions with 3 df

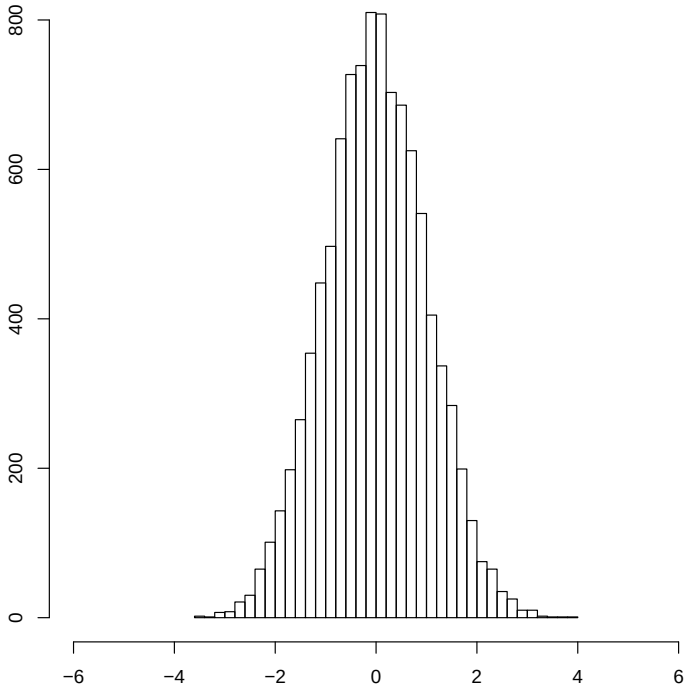


- n=10: First few observations: -0.69, 1.73, 0.55, 0.96, -0.79, $\cdots$
- n=100: Starting to look roughly "normal"
- n=1000: (12 obs outside plot. Range was: -12.8 to 8.3)
- n=10,000: (85 obs outside plot. Range was: -24.3 to 35.7)

Normal, t distributions: differences?

Normal distribution

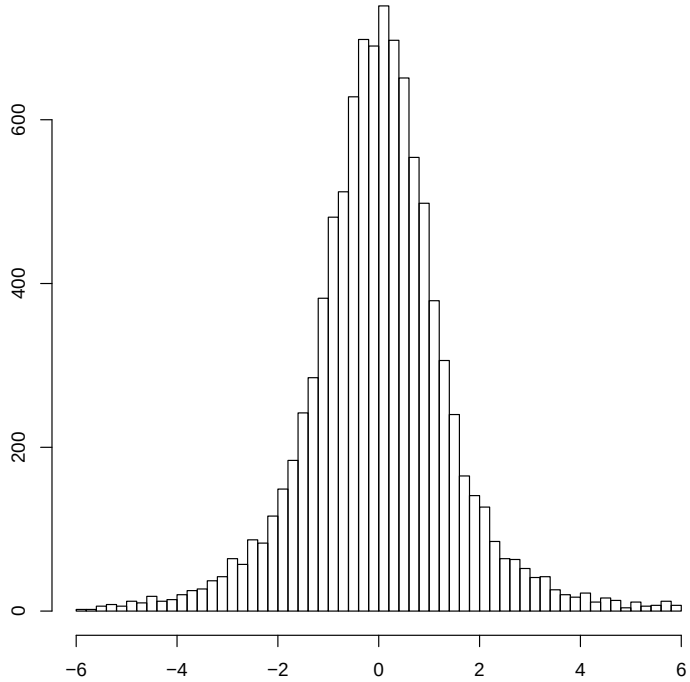
n= 10000



- Normal: range was: -3.6 to 3.9

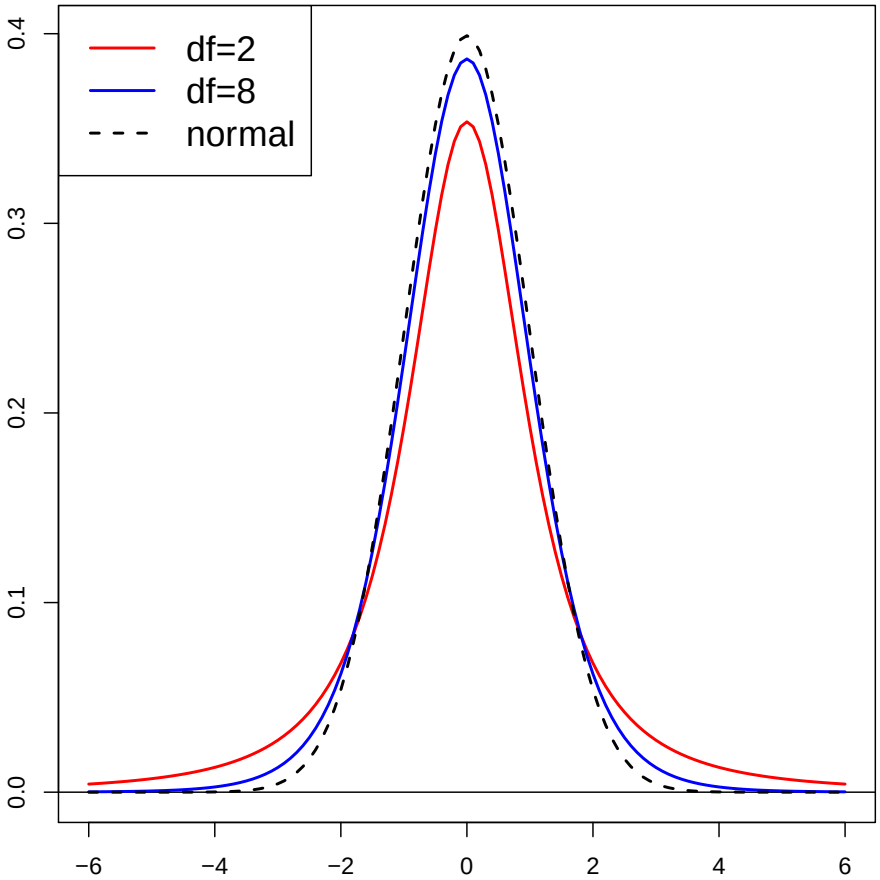
t_3 distribution

n= 10000



- t_3: range was: -24.3 to 35.7

Different t-distributions



With more data, larger n

- degrees of freedom (df) = $n - 1$ increases
- smaller area in tails of the distribution
- approaches normal distribution (which is t with ∞ df)

Normal, t distributions

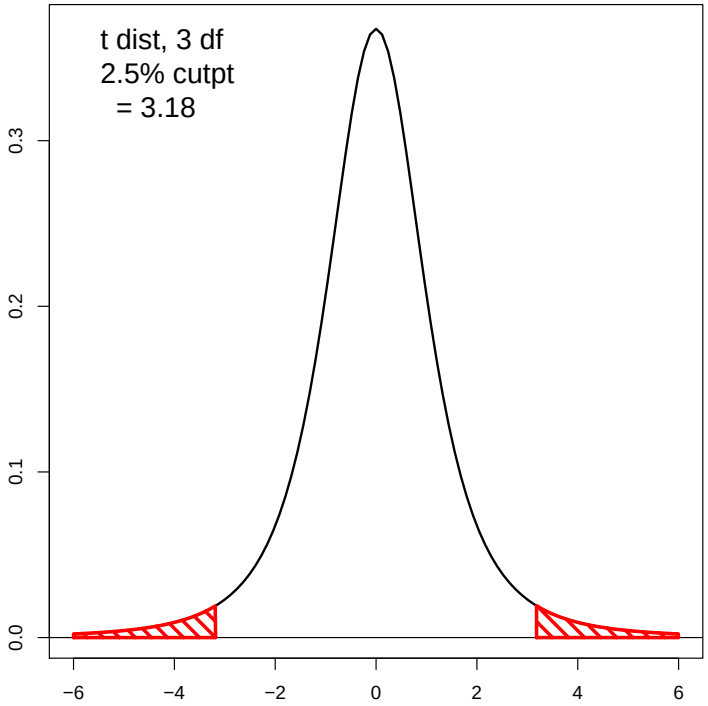
"Why doesn't $t_{calc} = \frac{\bar{x}-\mu_0}{SE(\bar{x})}$ have a normal distribution?"

Normal, t distributions

"Why doesn't $t_{calc} = \frac{\bar{x}-\mu_0}{SE(\bar{x})}$ have a normal distribution?"

- Recall we assumed $x_i \sim N(\mu, \sigma^2)$
- $\bar{x}$ has a normal distribution
- **If** we knew σ^2
  - SE($\bar{x}$) would be σ^2/n, a fixed number
  - "t_{calc}" would have a normal distribution
- When σ^2 not known, use s^2 to estimate it
  - Our estimate of σ^2 isn't perfect
  - Result: t_{calc} more variable

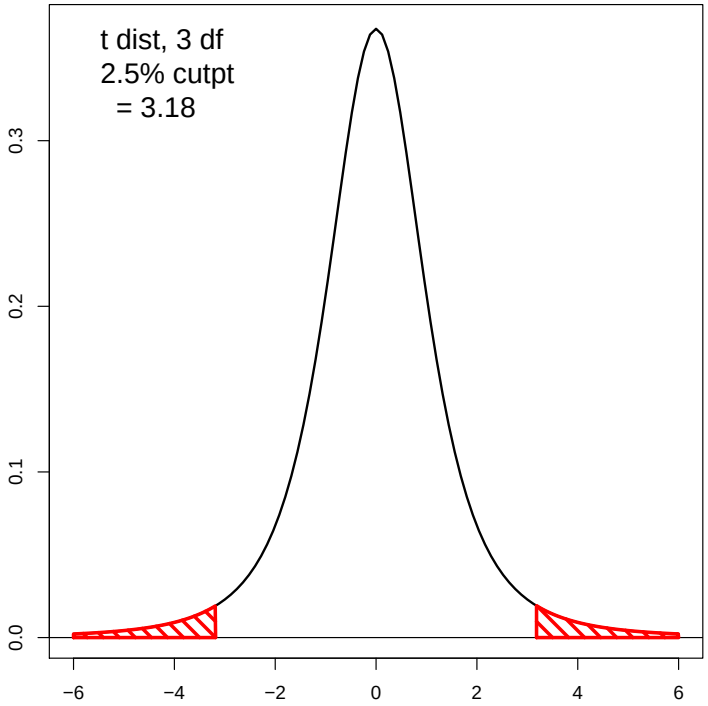
t-distribution (3 df)



If H_0 true

- Expect values near center
- Red: unusual

t-distribution (3 df)

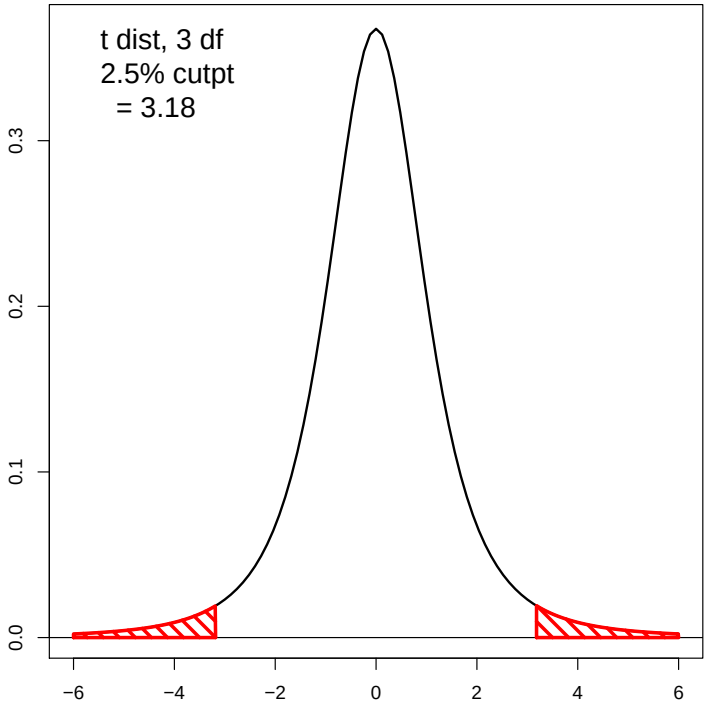


If H_0 true

- Expect values near center
- Red: unusual

- If H_0 true, control Pr(reject H_0) to α (α often chosen as 0.05)
- Reject H_0 if t_{calc} in shaded area
  - Each shaded area has 2.5% of the distribution
  - Shaded area defined by t_{crit}

t-distribution (3 df)

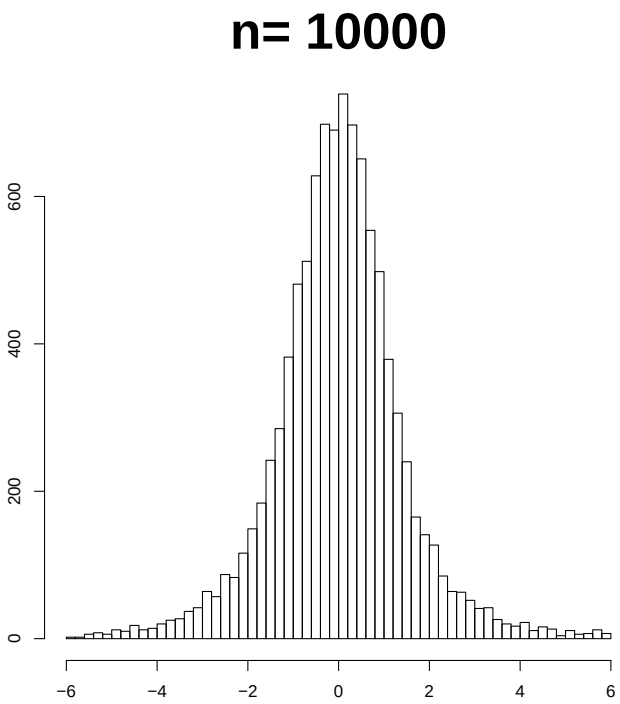


If H_0 true

- Expect values near center
- Red: unusual

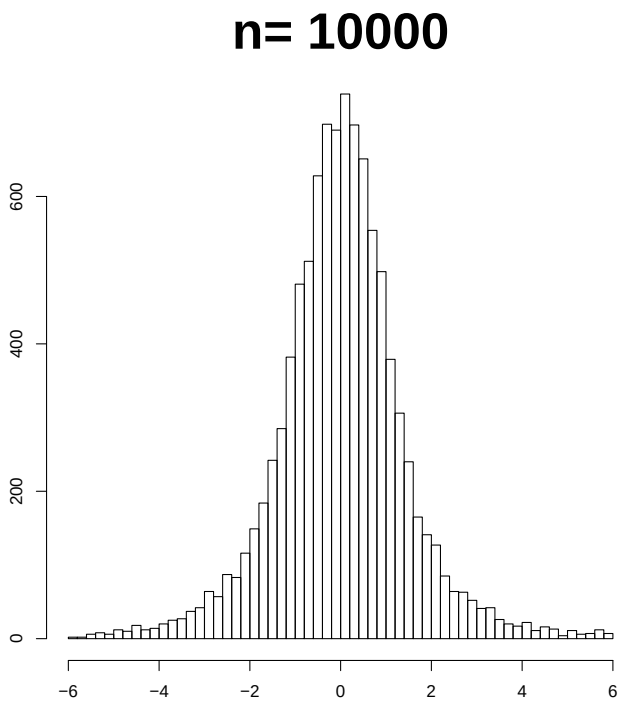
- If H_0 true, control Pr(reject H_0) to α (α often chosen as 0.05)
- Reject H_0 if t_{calc} in shaded area
  - Each shaded area has 2.5% of the distribution
  - Shaded area defined by t_{crit}
- t_{crit} is number for which 2.5% of distribution has larger values ($-t_{crit}$: 2.5% are smaller)
- t_{crit} depends on df and α
- With 3 df, $\alpha = 0.05$, $t_{crit} = 3.18$
- **Reject H_0 if** $|t_{calc}| > t_{crit}$

Sample from t-distribution: how good is t_{crit}?



- How good is t_{crit}?
- Do we *really* get 2.5% in each "tail"?
- For t_3, $\alpha = 0.05$, $t_{crit} = 3.18$.

Sample from t-distribution: how good is t_{crit}?



- How good is t_{crit}?
- Do we *really* get 2.5% in each "tail"?
- For t_3, $\alpha = 0.05$, $t_{crit} = 3.18$.

- From sample of 10,000 observation from t_3
  - 247 observations were < -3.18 (2.47%)
  - 258 observations were > 3.18 (2.58%)
- If H_0 true and $\alpha = 0.05$, approximately 5% of the time we would reject H_0 ("type I error")

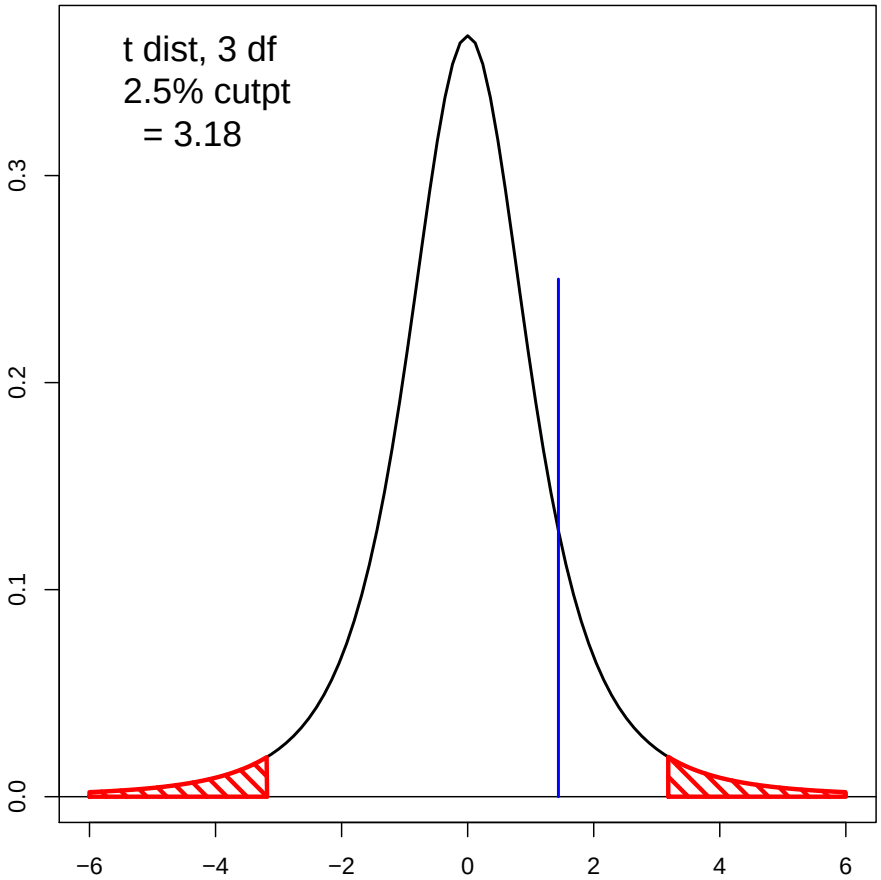
t-test for Dr. Forte

- Recall Dr. Forte's question: **is weight loss in C57BL/6 mice (his data) different than in BALB/c mice (published data, weight loss = 4 ounces)?**
- His hypotheses: $H_0 : \mu = 4, H_A : \mu \neq 4$.
- μ: weight loss in population of C57BL/6 mice
- Dr. Forte: $t_{calc} = \frac{\bar{x} - \mu_0}{SE(\bar{x})} = (5.5 - 4)/1.04 = 1.44$
- t-test rule: **Reject** H_0 **if** $|t_{calc}| > t_{crit}$
- With 3 df, $\alpha = 0.05$, $t_{crit} = 3.18$
- What should Dr. Forte conclude?

t-test for Dr. Forte

- Recall Dr. Forte's question: **is weight loss in C57BL/6 mice (his data) different than in BALB/c mice (published data, weight loss = 4 ounces)?**
- His hypotheses: $H_0 : \mu = 4, H_A : \mu \neq 4$.
- μ: weight loss in population of C57BL/6 mice
- Dr. Forte: $t_{calc} = \frac{\bar{x} - \mu_0}{SE(\bar{x})} = (5.5 - 4)/1.04 = 1.44$
- t-test rule: **Reject H_0 if** $|t_{calc}| > t_{crit}$
- With 3 df, $\alpha = 0.05$, $t_{crit} = 3.18$
- What should Dr. Forte conclude?
- Fail to reject H_0
  - Dr. Forte's data are consistent with H_0 ($p = 0.25$)
  - **No evidence that weight loss in C57BL/6 mice is different from 4 ounces.**

t-distribution (3df)



- Dr. Forte's value: blue line (1.44)
- Compare Dr. Forte's value to a *t* distribution with 3 df
- Reject H_0 if t_{calc} in shaded area
- Fail to reject H_0

Confidence interval

- 95% confidence interval (CI): $\bar{x} \pm t_{crit} \times \text{SE}(\bar{x})$
  - (Assumes t_{crit} calculated using $\alpha = 0.05$)
- This example: $t_{crit} = 3.18$
- We calculated $\bar{x} = 5.5, \text{SE}(\bar{x}) = 1.04$
- What is Dr. Forte's 95% CI?

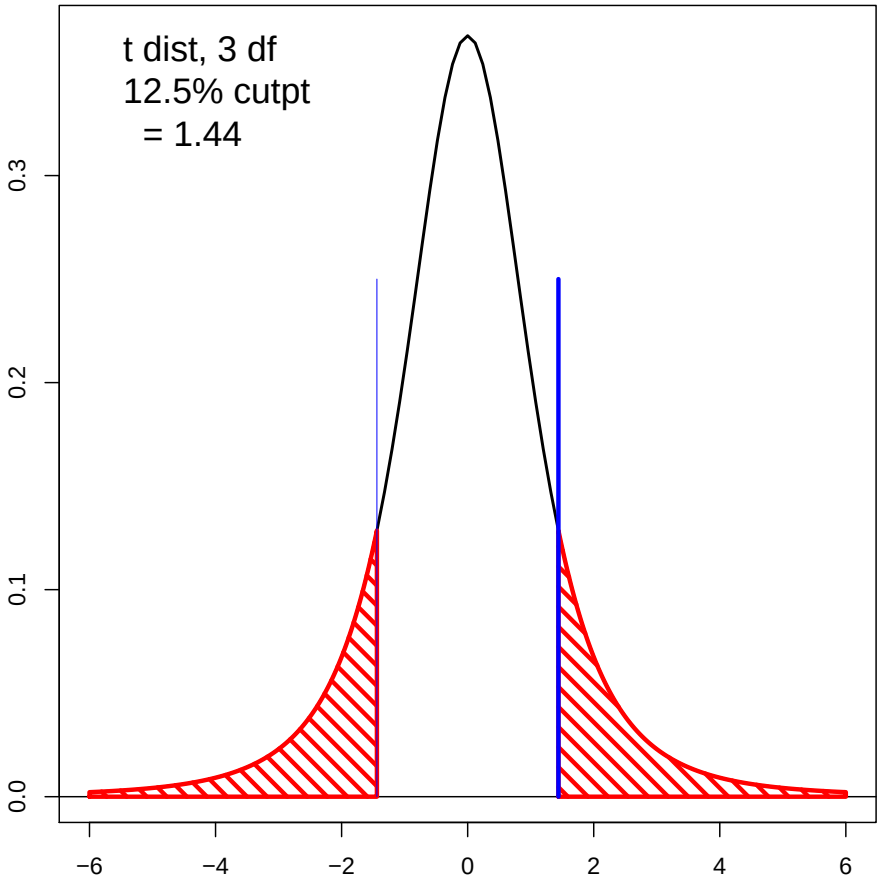
Confidence interval

- 95% confidence interval (CI): $\bar{x} \pm t_{crit} \times \text{SE}(\bar{x})$
  - (Assumes t_{crit} calculated using $\alpha = 0.05$)
- This example: $t_{crit} = 3.18$
- We calculated $\bar{x} = 5.5, \text{SE}(\bar{x}) = 1.04$
- What is Dr. Forte's 95% CI?
- 95% CI is $5.5 \pm 3.18 \times 1.04 = (2.19, 8.81)$
  - This is a 95% CI for μ
  - Fail to reject H_0 because interval includes $\mu_0 = 4$.
  - Confidence interval more informative than "fail to reject H_0"

Classical statistics

- μ considered fixed.
- Observed data (x_i) are considered random
  - Could collect new data: different x_i
  - Implies calculated confidence interval is random.
- If H_0 is true, on average 95% of the CIs will include μ.
- Can never answer "Is H_0 true?"
  - Can only say something about how unusual our test statistic is **if** H_0 is true

p-value



p-value = probability of observing a statistic as extreme or more extreme as we observed, if H_0 is true

- Curve: total area=1
- Dr. Forte's $t_{calc} = 1.44$.
- p-value = shaded area = 0.25

p-value

We calculated $p = 0.25$ for Dr. Forte's data.

- Is this the same as the probability that H_0 is true?

p-value

We calculated $p = 0.25$ for Dr. Forte's data.

- Is this the same as the probability that H_0 is true?

No

- p-value is calculated assuming H_0 is true.

Recall: p-value is probability of observing a statistic as or more extreme as we observed, **if** H_0 is true

p-value questions

What do you think of these statements?

- “The p-value tells you whether or not the observed effect is real.”

p-value questions

What do you think of these statements?

- “The p-value tells you whether or not the observed effect is real.” **Not true**

p-value questions

What do you think of these statements?

- “The p-value tells you whether or not the observed effect is real.” **Not true**
- “If the result is not statistically significant, that proves there is no difference.”

p-value questions

What do you think of these statements?

- “The p-value tells you whether or not the observed effect is real.” **Not true**
- “If the result is not statistically significant, that proves there is no difference.” **Not true**

p-value questions

What do you think of these statements?

- "The p-value tells you whether or not the observed effect is real." **Not true**
- "If the result is not statistically significant, that proves there is no difference." **Not true**

Better:

- Use knowledge of the magnitude of the estimated effect (and the CI) to help interpret the results.
- Suppose comparing treatment to control and $p = 0.07$
  - If estimated effect large, "Our study suggests an important benefit of treatment, but this did not reach statistical significance."
- A *p*-value is not an end in itself.

t-test for M&M's

- Consider M&M data from 1st sample of 8 people
  - Group 1: $\bar{x}$= 0.374, s_x= 0.0479 SE($\bar{x}$)= 0.0169
- Recall x_i is % (red + blue) M&M's
- Test $H_0 : \mu = 1/3$, versus $H_A : \mu \neq 1/3$.
- What is t_{calc}?

t-test for M&M's

- Consider M&M data from 1st sample of 8 people
  - Group 1: $\bar{x}$= 0.374, s_x= 0.0479 SE($\bar{x}$)= 0.0169
- Recall x_i is % (red + blue) M&M's
- Test $H_0 : \mu = 1/3$, versus $H_A : \mu \neq 1/3$.
- What is t_{calc}?

$$\frac{0.374 - 0.333}{0.0169} = 2.41$$

- Compare to $t_{crit} = 2.36$ (based on (n-1) df, $\alpha = 0.05$)
- Conclusion?

t-test for M&M's

- Consider M&M data from 1st sample of 8 people
  - Group 1: $\bar{x}$= 0.374, s_x= 0.0479 SE($\bar{x}$)= 0.0169
- Recall x_i is % (red + blue) M&M's
- Test $H_0 : \mu = 1/3$, versus $H_A : \mu \neq 1/3$.
- What is t_{calc}?

$$\frac{0.374 - 0.333}{0.0169} = 2.41$$

- Compare to $t_{crit} = 2.36$ (based on (n-1) df, $\alpha = 0.05$)
- Conclusion? Reject H_0 (also $p = 0.047$).
  - For first 8 people, mean % (red + blue) M&M's significantly different than 1/3.

t-test for M&M's

For groups of 8 people, $t_{crit} = 2.36$

- Group 1: $\bar{x}$= 0.374, s_x= 0.0479 SE($\bar{x}$)= 0.0169
  - $t_{calc} = 2.41$, $p = 0.047$
- Group 2: $\bar{x}$= 0.3564, s_x= 0.0574 SE($\bar{x}$)= 0.0203
  - $t_{calc} = 1.138$, $p = 0.293$
- Group 3: $\bar{x}$= 0.3549, s_x= 0.0348 SE($\bar{x}$)= 0.0123
  - $t_{calc} = 1.175$, $p = 0.123$
- Group 4: $\bar{x}$= 0.3465, s_x= 0.0342 SE($\bar{x}$)= 0.0121
  - $t_{calc} = 1.084$, $p = 0.314$

t-test for M&M's

For groups of 8 people, $t_{crit} = 2.36$

- Group 1: $\bar{x}$= 0.374, s_x= 0.0479 SE($\bar{x}$)= 0.0169
  - $t_{calc} = 2.41$, $p = 0.047$
- Group 2: $\bar{x}$= 0.3564, s_x= 0.0574 SE($\bar{x}$)= 0.0203
  - $t_{calc} = 1.138$, $p = 0.293$
- Group 3: $\bar{x}$= 0.3549, s_x= 0.0348 SE($\bar{x}$)= 0.0123
  - $t_{calc} = 1.175$, $p = 0.123$
- Group 4: $\bar{x}$= 0.3465, s_x= 0.0342 SE($\bar{x}$)= 0.0121
  - $t_{calc} = 1.084$, $p = 0.314$

For entire sample (n=32), $t_{crit} = 2.04$

- Everyone: $\bar{x}$= 0.358 s_x= 0.0436 SE($\bar{x}$)= 0.0077
  - $t_{calc} = 3.19$, $p = 0.003$

t-test for M&M's

For groups of 8 people, $t_{crit} = 2.36$

- Group 1: $\bar{x}$= 0.374, s_x= 0.0479 SE($\bar{x}$)= 0.0169
  - $t_{calc} = 2.41$, $p = 0.047$
- Group 2: $\bar{x}$= 0.3564, s_x= 0.0574 SE($\bar{x}$)= 0.0203
  - $t_{calc} = 1.138$, $p = 0.293$
- Group 3: $\bar{x}$= 0.3549, s_x= 0.0348 SE($\bar{x}$)= 0.0123
  - $t_{calc} = 1.175$, $p = 0.123$
- Group 4: $\bar{x}$= 0.3465, s_x= 0.0342 SE($\bar{x}$)= 0.0121
  - $t_{calc} = 1.084$, $p = 0.314$

For entire sample (n=32), $t_{crit} = 2.04$

- Everyone: $\bar{x}$= 0.358 s_x= 0.0436 SE($\bar{x}$)= 0.0077
  - $t_{calc} = 3.19$, $p = 0.003$

Larger n: smaller SE($\bar{x}$), smaller t_{crit}, easier to reject H_0 if H_0 not true.

Outline

2-sample *t*-test

- Dr. Forte now wants to compare the effect of 'superdrug' on weight loss in knockout mice (new) versus C57BL/6 mice (just measured)
  - NOTE: Better to collect data on both species in the same time period.
- Let $\bar{x}_1$ and s_1 be mean and SD of weight loss in 1st sample (C57BL/6), $n_1 = 4$ observations.
- Let $\bar{x}_2$ and s_2 be mean and SD of weight loss in 2nd sample (knockout), n_2 observations
- μ_1 is population mean weight loss in C57BL/6 mice, μ_2 in knockout mice.
- Null hypothesis of "no effect" or "no difference" is $H_0 : \mu_1 = \mu_2$, so $H_A : \mu_1 \neq \mu_2$.

2-sample t-test (cont.)

- $H_0 : \mu_1 = \mu_2$, $H_A : \mu_1 \neq \mu_2$
- Same as $H_0 : \mu_1 - \mu_2 = 0$ and $H_A : \mu_1 - \mu_2 \neq 0$.
- Estimate $\mu_1 - \mu_2$ by $\bar{x}_1 - \bar{x}_2$.
- $t_{calc} = \frac{\bar{x}_1 - \bar{x}_2}{SE(\bar{x}_1 - \bar{x}_2)}$

Assumptions of the 2-sample test: data from 2 populations

- Are independent
- Are normally distributed (within population)
- Have the same variance (or: use modified estimate of standard error)

Reject H_0 if $|t_{calc}| > t_{crit}$ (t_{crit} depends on $n_1 + n_2$ and α)

Scenario 1

- Pairs of mice are housed in the same cage with same food source
- Interest is in effect of restricted food intake
- Mice fed 2x/day, only 1 mouse can eat at once

Scenario 1

- Pairs of mice are housed in the same cage with same food source
- Interest is in effect of restricted food intake
- Mice fed 2x/day, only 1 mouse can eat at once

Is this a reasonable design?

Scenario 1

- Pairs of mice are housed in the same cage with same food source
- Interest is in effect of restricted food intake
- Mice fed 2x/day, only 1 mouse can eat at once

Is this a reasonable design?

Mice are not independent: if one mouse in a pair eats aggressively, the other mouse cannot eat as much

Scenario 2

- 12 mice receive standard food, 12 a new food.
- Interest: gain in head circumference, measured at end of 2 weeks
- Bob measures mice receiving standard food, Harry measures the others
- Bob and Harry do not know which mice received which food source

Scenario 2

- 12 mice receive standard food, 12 a new food.
- Interest: gain in head circumference, measured at end of 2 weeks
- Bob measures mice receiving standard food, Harry measures the others
- Bob and Harry do not know which mice received which food source

Is this a reasonable design?

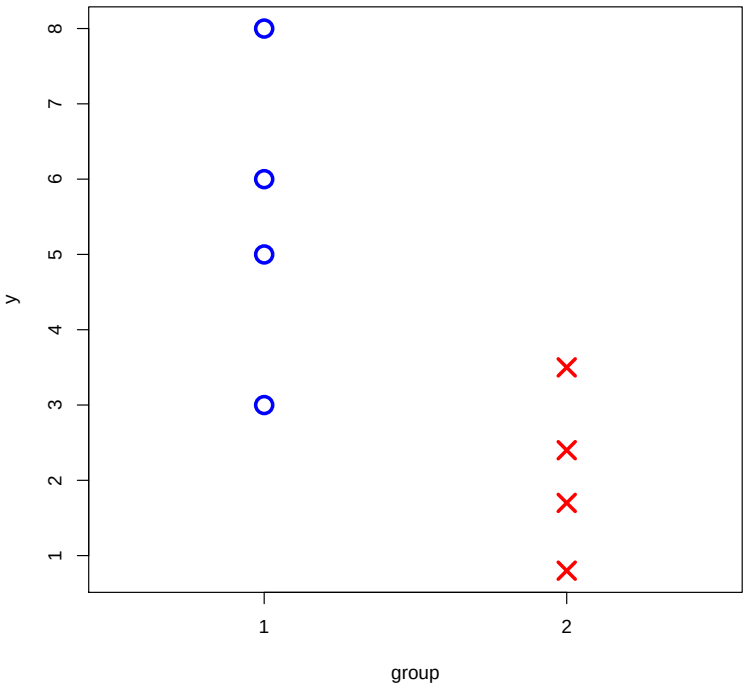
Scenario 2

- 12 mice receive standard food, 12 a new food.
- Interest: gain in head circumference, measured at end of 2 weeks
- Bob measures mice receiving standard food, Harry measures the others
- Bob and Harry do not know which mice received which food source

Is this a reasonable design? **No**

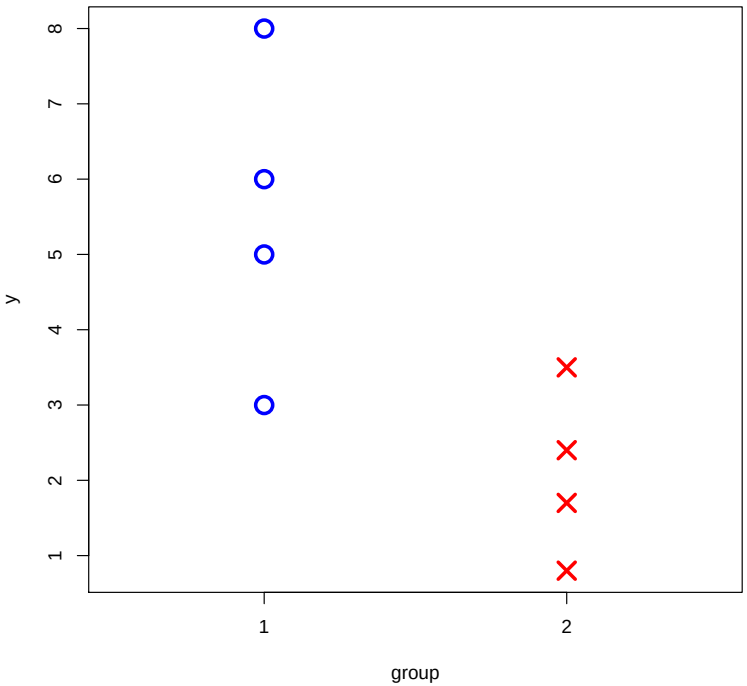
- Food source is totally confounded with person who takes measurements
- Bob might tend to get smaller measurements on same animal (bias)
- Harry might tend to measure less accurately (larger variance)

Dr. Forte's 2-sample data



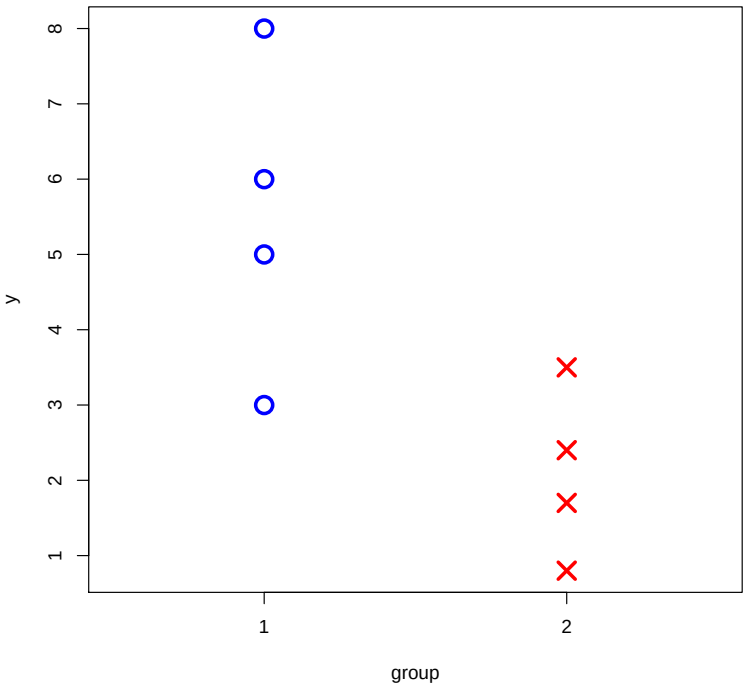
- Do you think assumptions are reasonable?

Dr. Forte's 2-sample data



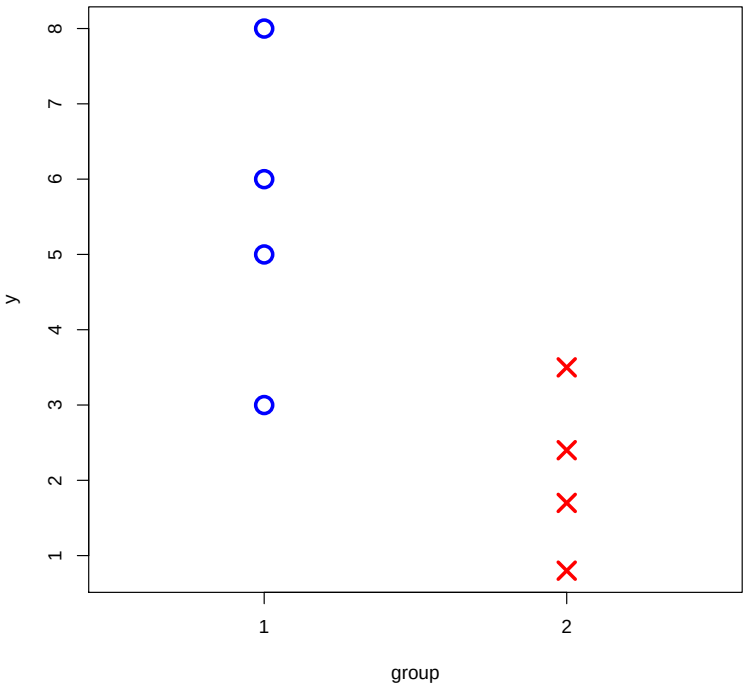
- Do you think assumptions are reasonable? (OK)
- $H_0 : \mu_1 - \mu_2 = 0$, $H_A : \mu_1 - \mu_2 \neq 0$.
- $t_{calc} = \frac{\bar{x}_1 - \bar{x}_2}{SE(\bar{x}_1 - \bar{x}_2)}$
- $\bar{x}_1 = 5.5, \bar{x}_2 = 2.1$
- $SE(\bar{x}_1 - \bar{x}_2) = 1.19$ (assumes same variance in 2 groups)
- What is t_{calc}?

Dr. Forte's 2-sample data



- Do you think assumptions are reasonable? (OK)
- $H_0 : \mu_1 - \mu_2 = 0$, $H_A : \mu_1 - \mu_2 \neq 0$.
- $t_{calc} = \frac{\bar{x}_1 - \bar{x}_2}{SE(\bar{x}_1 - \bar{x}_2)}$
- $\bar{x}_1 = 5.5, \bar{x}_2 = 2.1$
- $SE(\bar{x}_1 - \bar{x}_2) = 1.19$ (assumes same variance in 2 groups)
- What is t_{calc}? $\frac{5.5-2.1}{1.19} = 2.86$
- $t_{crit} = 2.45$, from 6 df, $\alpha = 0.05$
- What is your conclusion?

Dr. Forte's 2-sample data



- Do you think assumptions are reasonable? (OK)
- $H_0 : \mu_1 - \mu_2 = 0$, $H_A : \mu_1 - \mu_2 \neq 0$.
- $t_{calc} = \frac{\bar{x}_1 - \bar{x}_2}{SE(\bar{x}_1 - \bar{x}_2)}$
- $\bar{x}_1 = 5.5, \bar{x}_2 = 2.1$
- $SE(\bar{x}_1 - \bar{x}_2) = 1.19$ (assumes same variance in 2 groups)
- What is t_{calc}? $\frac{5.5-2.1}{1.19} = 2.86$
- $t_{crit} = 2.45$, from 6 df, $\alpha = 0.05$
- What is your conclusion? (Reject H_0)

95% CI for $\mu_1 - \mu_2$ is $\bar{x}_1 - \bar{x}_2 \pm t_{crit} \times \text{SE}(\bar{x}_1 - \bar{x}_2) = (0.50, 6.30)$

Dr. Forte's conclusions

- Recall: **Dr. Forte wants to compare the effect of 'superdrug' on weight loss in knockout vs C57BL/6 mice.**
- C57BL/6 mice: ($\bar{x}_1 = 5.5$), knockout mice ($\bar{x}_2 = 2.1$)
- We rejected H_0, $p = 0.03$

What does "reject H_0" mean here?

Dr. Forte's conclusions

- Recall: **Dr. Forte wants to compare the effect of 'superdrug' on weight loss in knockout vs C57BL/6 mice.**
- C57BL/6 mice: ($\bar{x}_1 = 5.5$), knockout mice ($\bar{x}_2 = 2.1$)
- We rejected H_0, $p = 0.03$

What does "reject H_0" mean here? **Data suggests that the average weight loss in C57BL/6 mice is significantly greater than in knockout mice.**

What does the p-value mean?

Dr. Forte's conclusions

- Recall: **Dr. Forte wants to compare the effect of 'superdrug' on weight loss in knockout vs C57BL/6 mice.**
- C57BL/6 mice: ($\bar{x}_1 = 5.5$), knockout mice ($\bar{x}_2 = 2.1$)
- We rejected H_0, $p = 0.03$

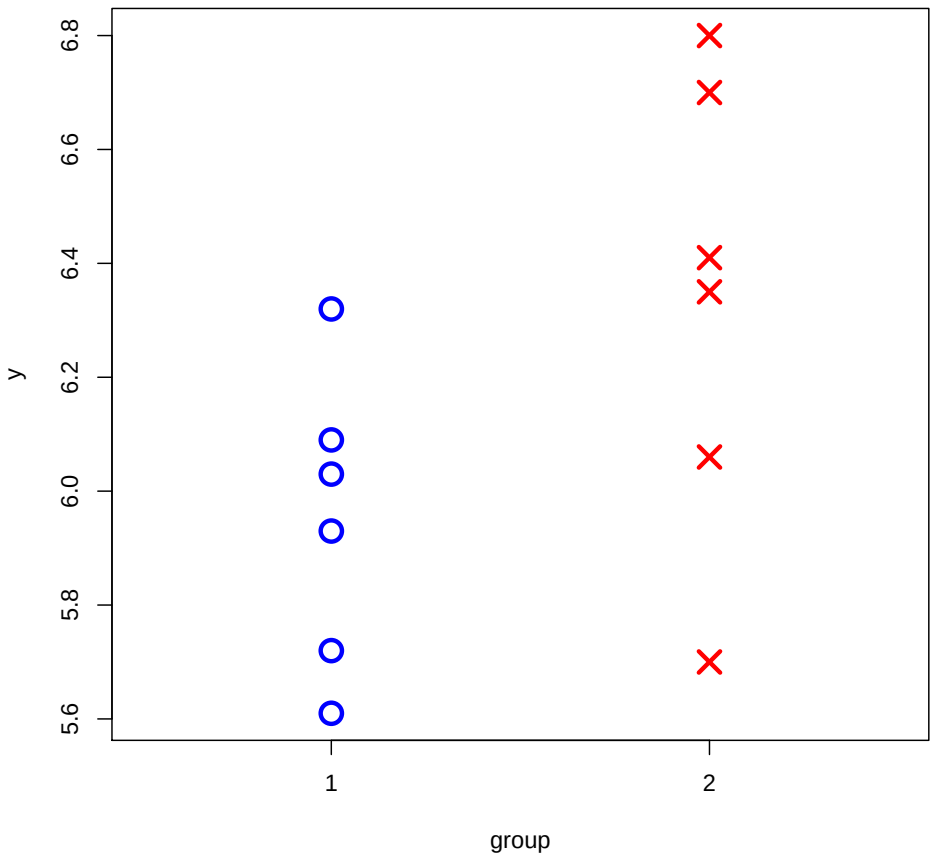
What does "reject H_0" mean here? **Data suggests that the average weight loss in C57BL/6 mice is significantly greater than in knockout mice.**

What does the p-value mean?

- If the 2 species had the same mean weight loss, we would have observed a difference as great or greater 3% of the time.

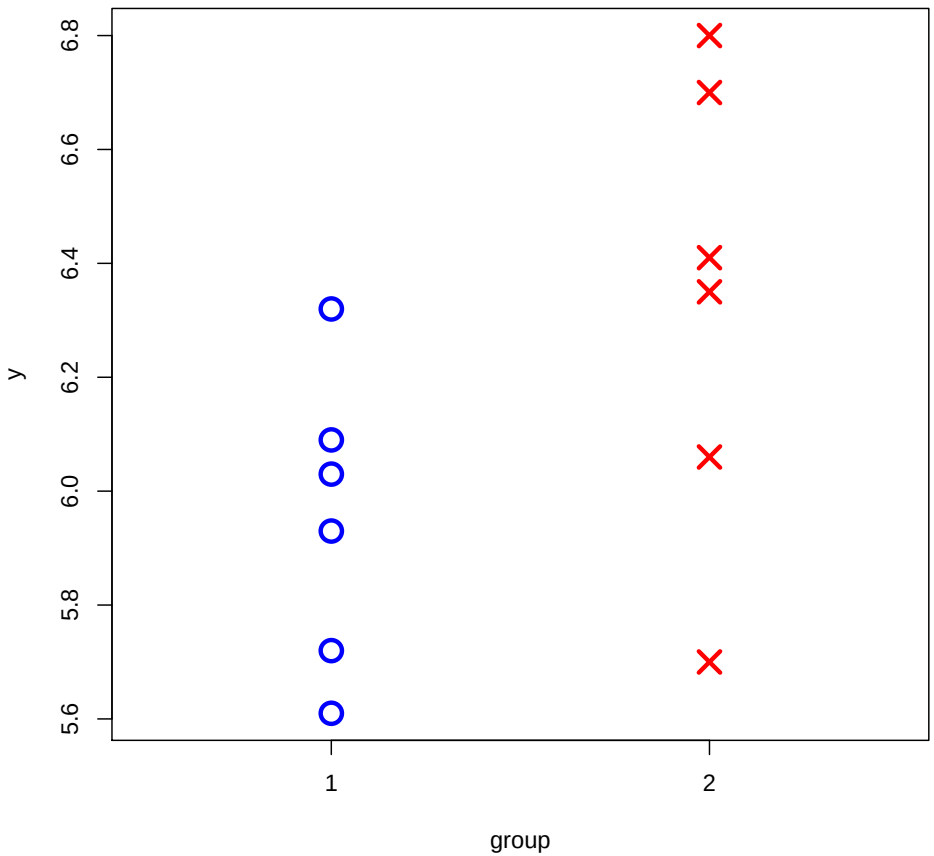
Outline

Plot the data: why?



Comments?

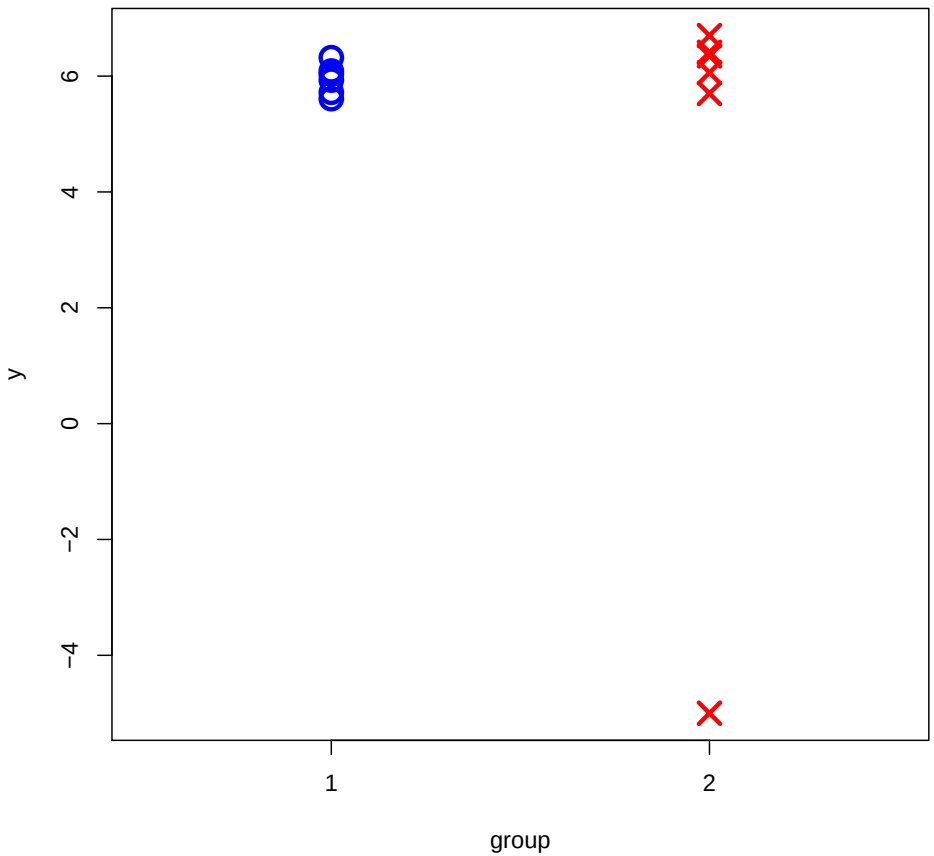
Plot the data: why?



Comments?

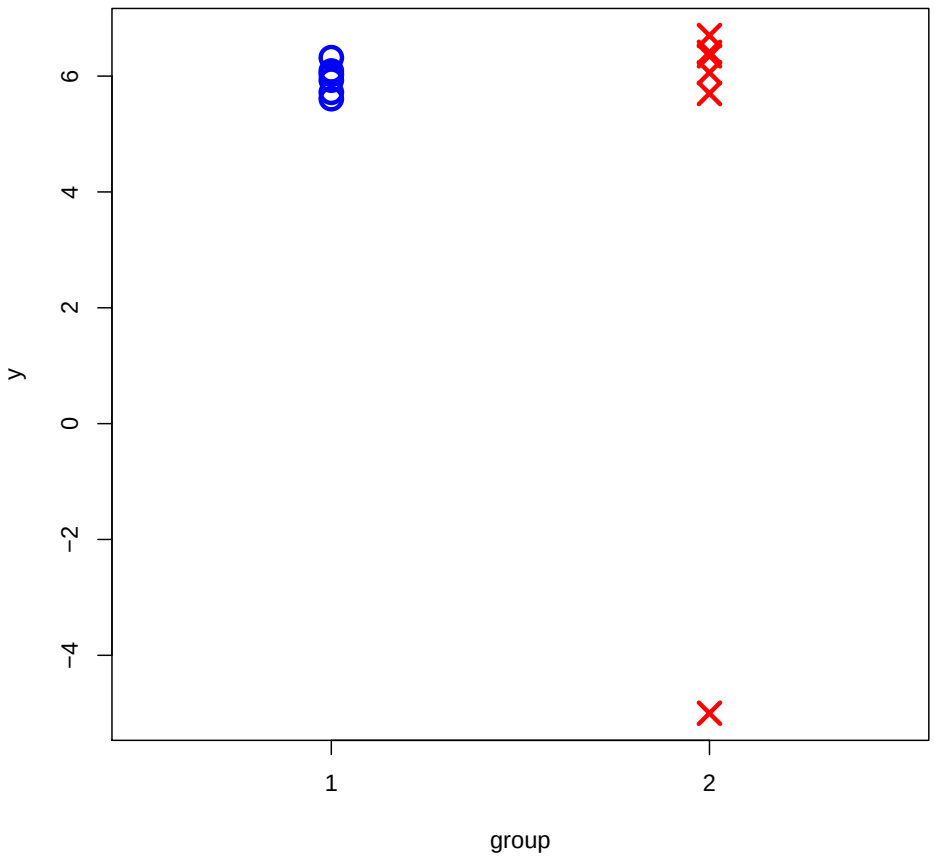
- The data distribution looks reasonable

Plot the data: why?



Comments?

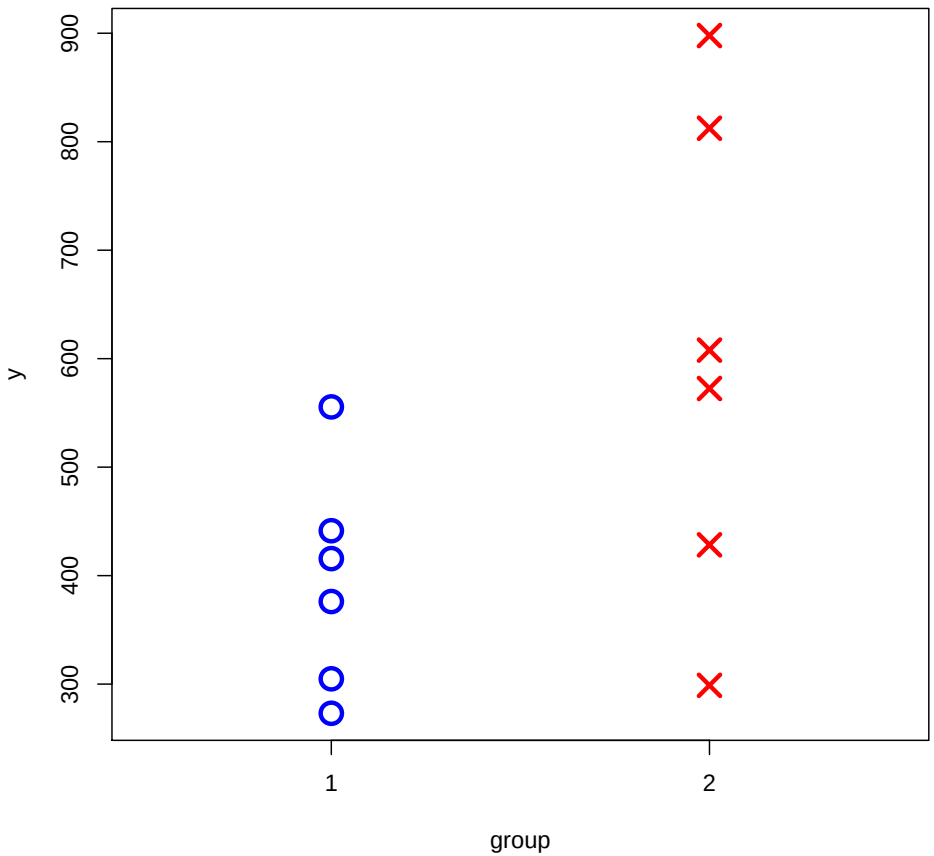
Plot the data: why?



Comments?
One large outlying observation

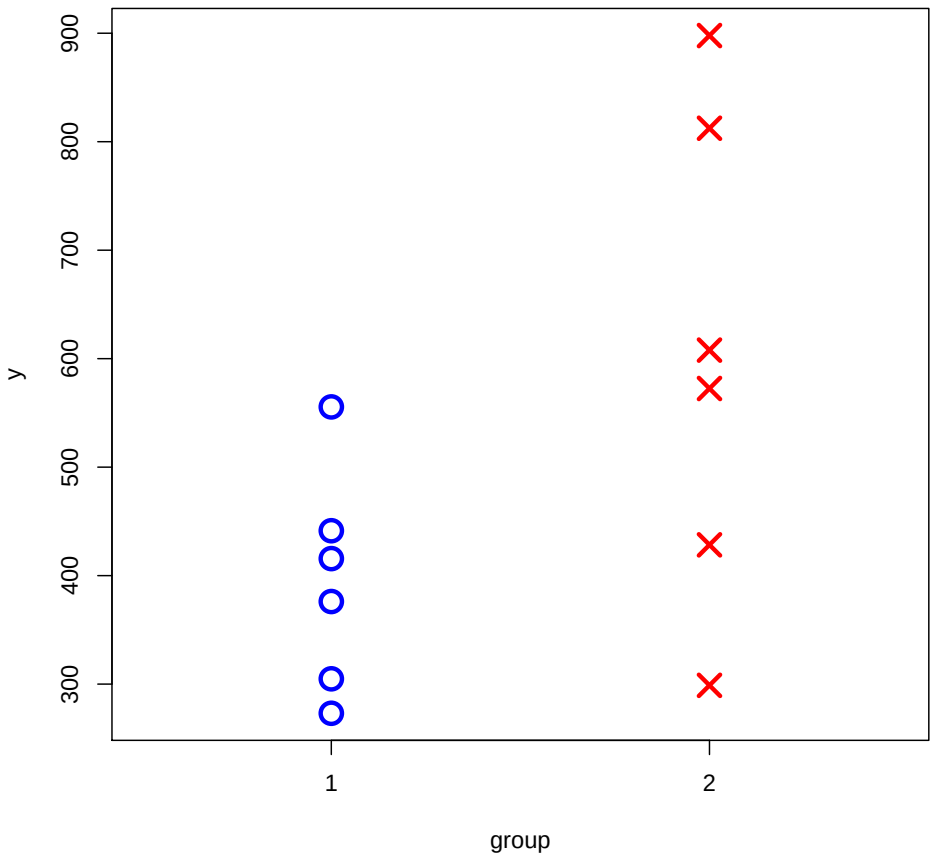
- A mistake?
- Non-normal distribution?

Plot the data: why?



Comments?

Plot the data: why?



Comments?

- Variability within groups not similar
- Same data as first plot - EXCEPT data is exponentiated
- Suggest: use log transformation
- Knowledge of what the outcomes are can guide decision about whether to transform

Outline

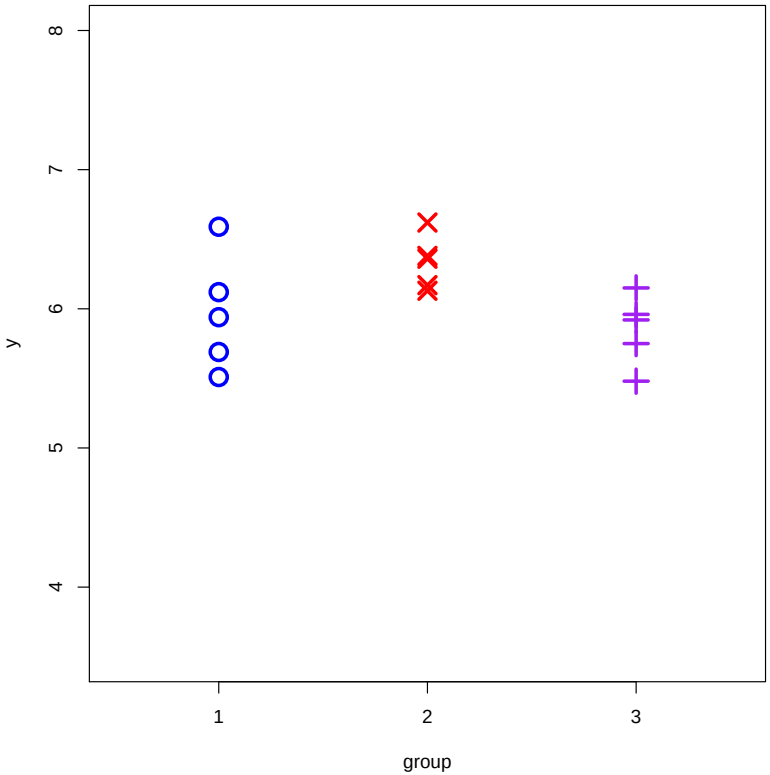
Dr. Crescendo

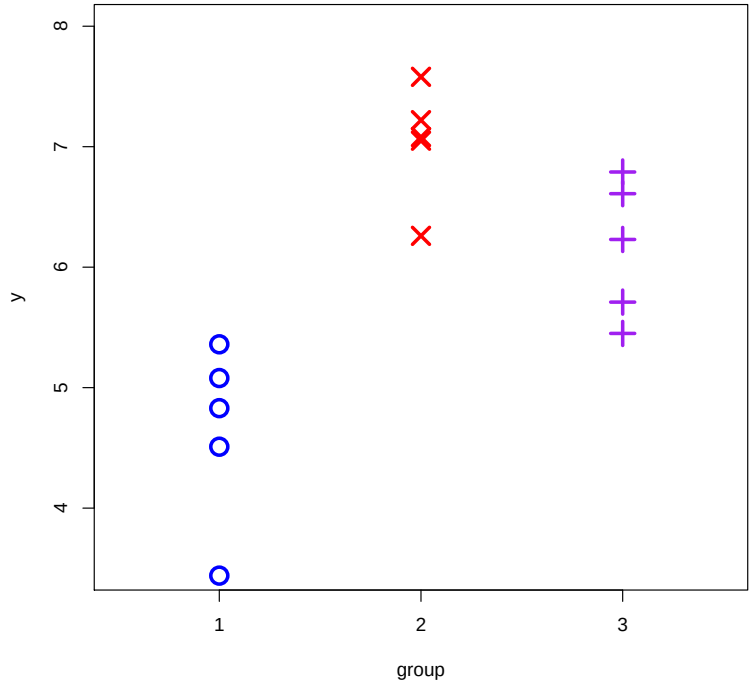
- Dr. Crescendo wants to test the effects of drugs A, B, and C on cell counts in 96-well plates
- Some things to consider
  - Is each drug tested on multiple plates? (plate effect)
  - Are the wells at the edge of the plate different from others? (edge effect)
  - Are the wells for drug A always read first? (order effect)

Set up for 1-way ANOVA

- Dr. Crescendo took 5 measurements from each of the 3 treatments (A,B,C)
- μ_1, μ_2, μ_3 are population means of treatment effects A, B, and C.
- Hypotheses:
  - $H_0 : \mu_1 = \mu_2 = \mu_3$
  - H_A : at least 2 means are not equal.
- How to test H_0, H_A?

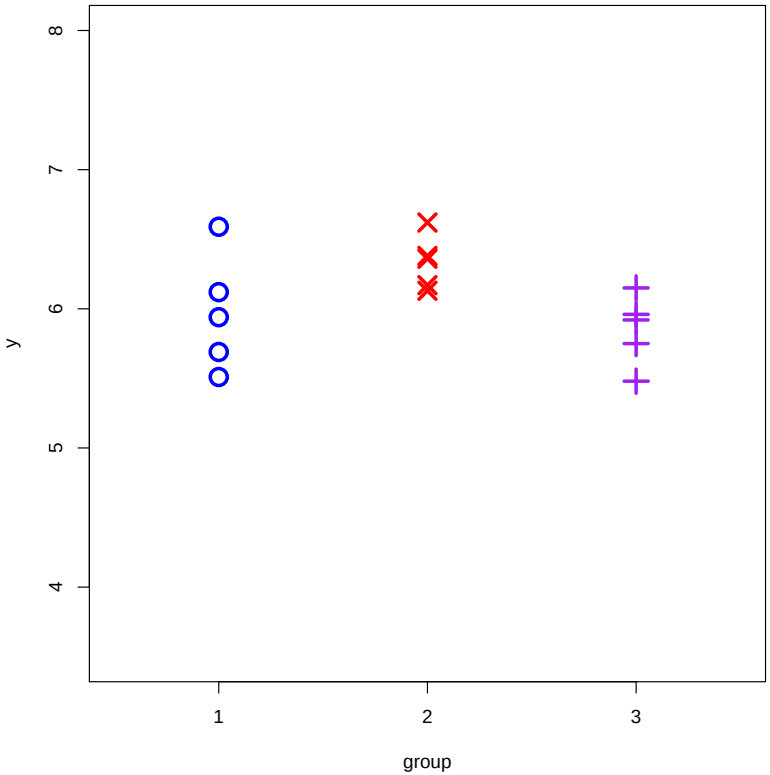
Pictures for 1-way ANOVA

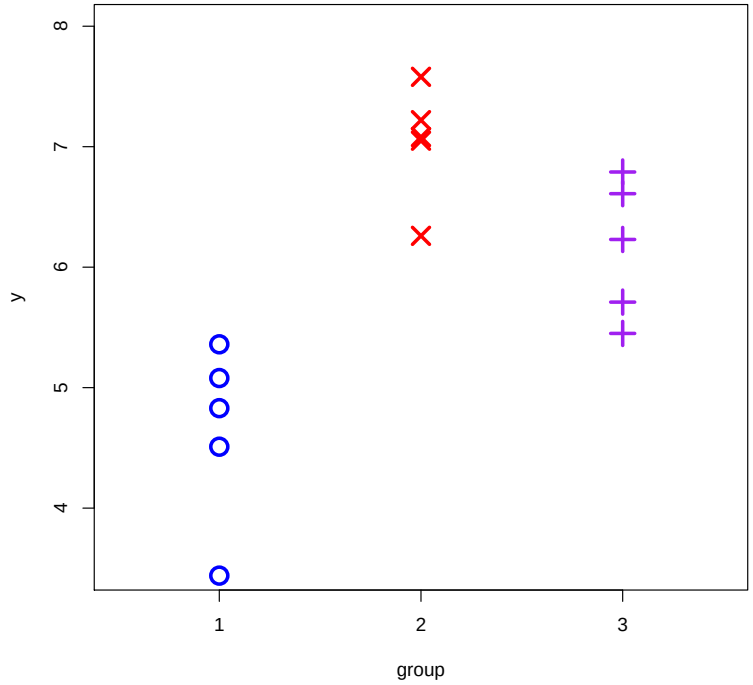




How do the 2 pictures differ?

Pictures for 1-way ANOVA

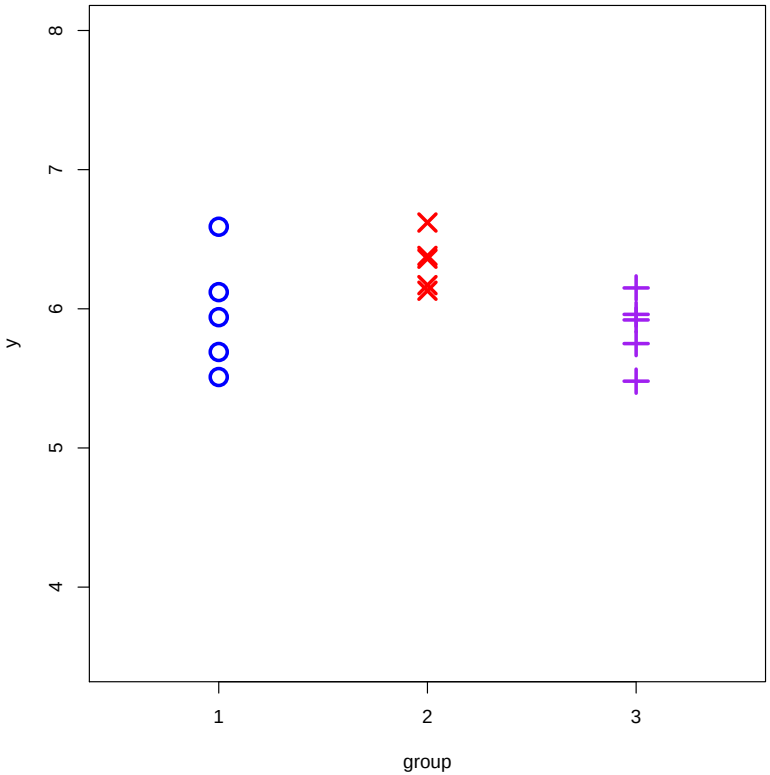


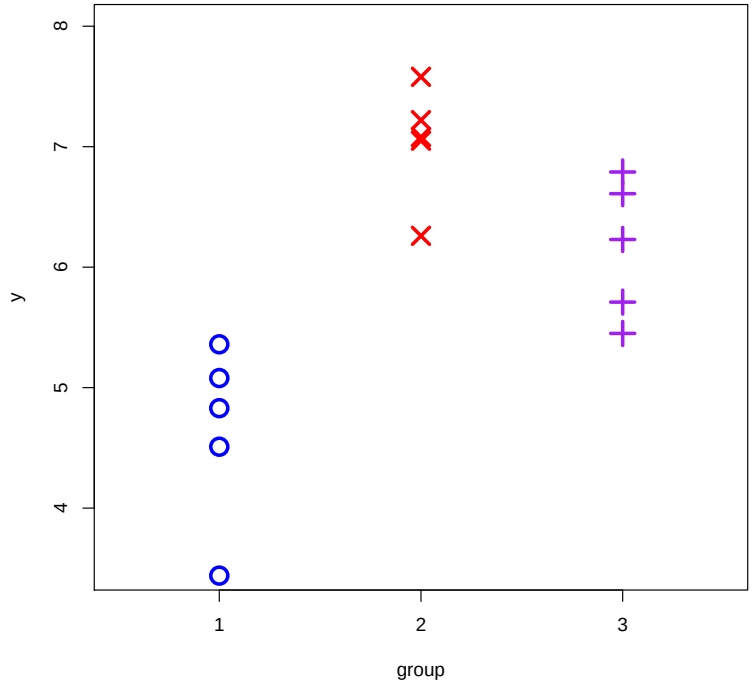


How do the 2 pictures differ?

- Right: 3 means are more different from each other

Pictures for 1-way ANOVA

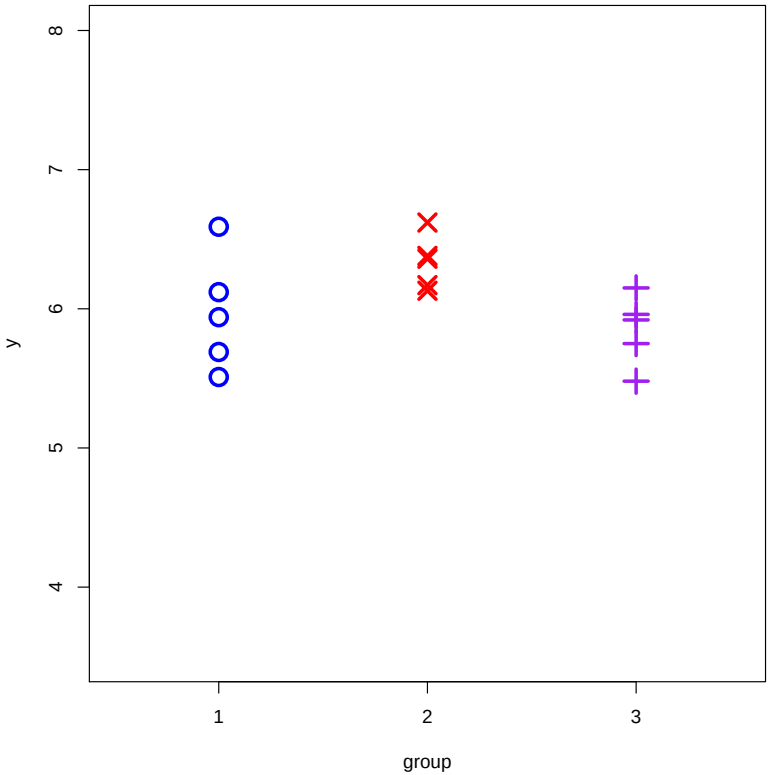


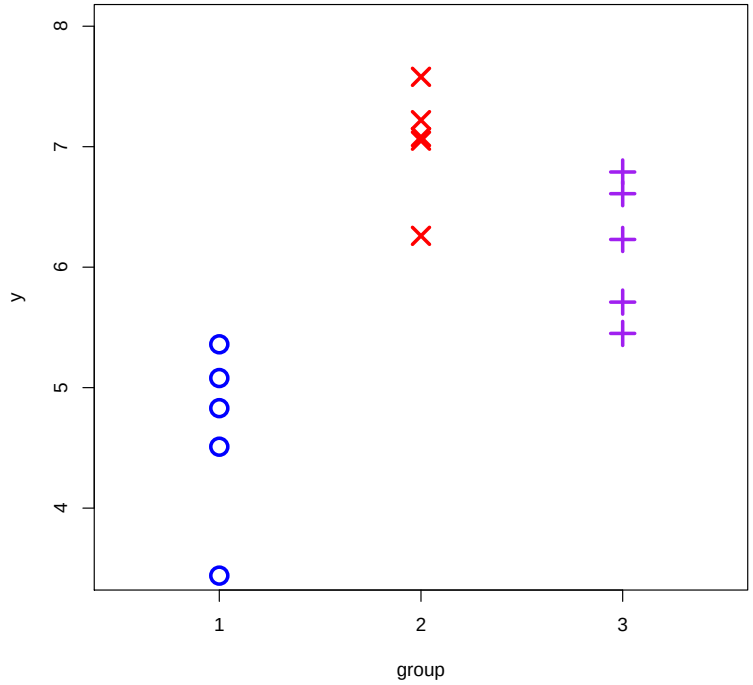


How do the 2 pictures differ?

- Right: 3 means are more different from each other
- Right: Variability **between** means is also more different

Pictures for 1-way ANOVA

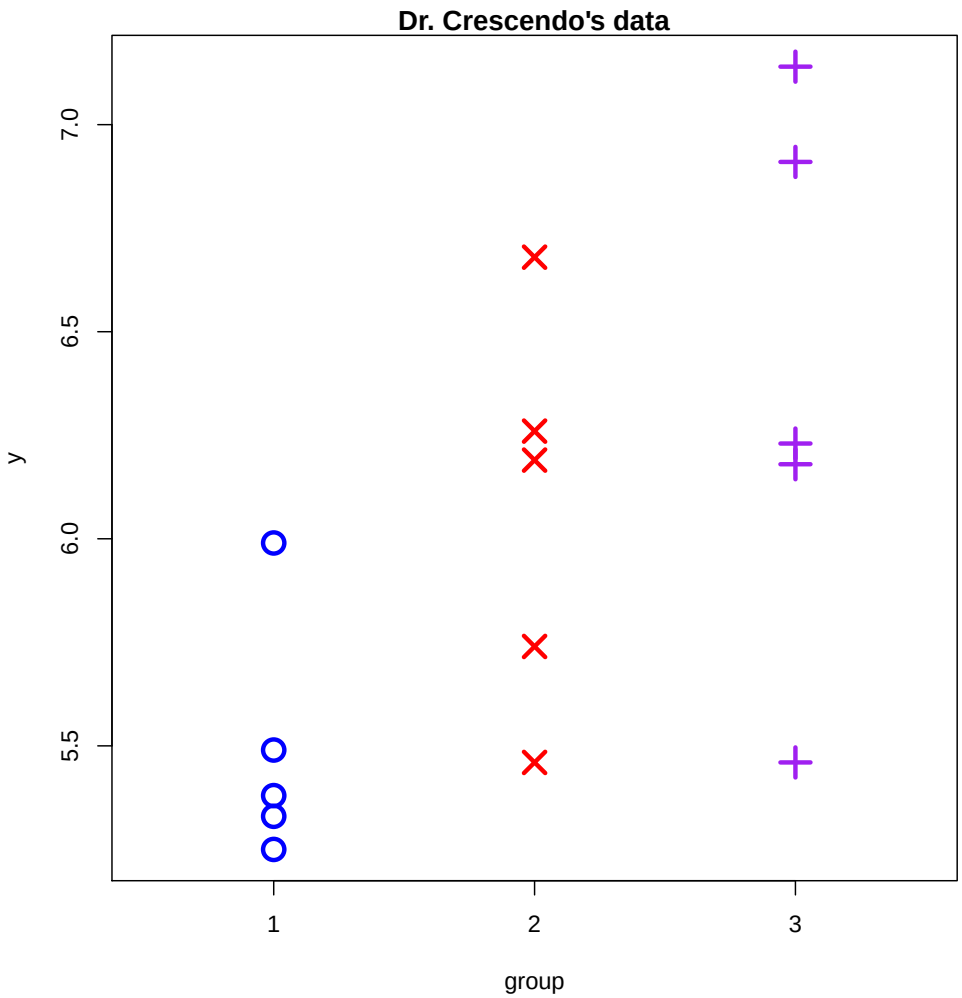




How do the 2 pictures differ?

- Right: 3 means are more different from each other
- Right: Variability **between** means is also more different
- ANOVA compares variability between treatments to variability within treatments

Dr Crescendo's data



- Sample sizes: $n_1 = n_2 = n_3 = 5$.
- Sample means: $\bar{x}_1 = 5.49, \bar{x}_2 = 6.07, \bar{x}_3 = 6.38$.
- Sample variances: $s_1^2 = 0.29, s_2^2 = 0.48, s_3^2 = 0.66$

1-way ANOVA: basics

- ANOVA assumes $\text{Var}(x_1) = \text{Var}(x_2) = \text{Var}(x_3)$
  - Call this σ^2

1-way ANOVA: basics

- ANOVA assumes $\text{Var}(x_1) = \text{Var}(x_2) = \text{Var}(x_3)$
  - Call this σ^2
- Variability within treatments is weighted average of s_1^2, s_2^2, s_3^2.
  - degrees of freedom (df) = $n_1 + n_2 + n_3 - 3$ (here: df=12)
  - Dr. Crescendo: within treatment variance is $3.016/12 = 0.25$

1-way ANOVA: basics

- ANOVA assumes $\text{Var}(x_1) = \text{Var}(x_2) = \text{Var}(x_3)$
  - Call this σ^2
- Variability within treatments is weighted average of s_1^2, s_2^2, s_3^2.
  - degrees of freedom (df) = $n_1 + n_2 + n_3 - 3$ (here: df=12)
  - Dr. Crescendo: within treatment variance is $3.016/12 = 0.25$
- Variability between treatments is weighted average of how far apart group means are from overall mean
  - df = number groups - 1 (here: df=2)
  - Dr. Crescendo: between treatment variance is $2.063/2 = 1.03$

Testing in 1-way ANOVA (3 groups)

- Recall: 1-way ANOVA compares
  - Variability within treatment: MS(Within) = MS(Error)
  - Variability between treatment: MS(Between) = MS(Group)
  - (Note: MS = mean square)
- Dr. Crescendo:
  - MS(Error) $= \frac{3.016}{12} = 0.25$
  - MS(Between) $= \frac{2.063}{2} = 1.03$

Testing in 1-way ANOVA (3 groups)

- Recall: 1-way ANOVA compares
  - Variability within treatment: MS(Within) = MS(Error)
  - Variability between treatment: MS(Between) = MS(Group)
  - (Note: MS = mean square)
- Dr. Crescendo:
  - MS(Error) $= \frac{3.016}{12} = 0.25$
  - MS(Between) $= \frac{2.063}{2} = 1.03$

ANOVA table

	Df	Sum Sq	Mean Sq
Between	2	2.06337	1.03169
Error	12	3.01612	0.25134

1-way ANOVA: basics

If H_0 true

- How do MS(Within) and MS(Between) compare?
- How could we estimate σ^2?

1-way ANOVA: basics

If H_0 true

- How do MS(Within) and MS(Between) compare?
- How could we estimate σ^2?
- MS(Between) $\approx$ MS(Within), and both estimate σ^2
- $\frac{\text{MS(Between)}}{\text{MS(Within)}}$ close to 1.

1-way ANOVA: basics

If H_0 true

- How do MS(Within) and MS(Between) compare?
- How could we estimate σ^2?
- MS(Between) $\approx$ MS(Within), and both estimate σ^2
- $\frac{\text{MS(Between)}}{\text{MS(Within)}}$ close to 1.

If H_0 not true, same questions

1-way ANOVA: basics

If H_0 true

- How do MS(Within) and MS(Between) compare?
- How could we estimate σ^2?
- MS(Between) $\approx$ MS(Within), and both estimate σ^2
- $\frac{\text{MS(Between)}}{\text{MS(Within)}}$ close to 1.

If H_0 not true, same questions

- **Only** MS(Within) estimates σ^2.
- MS(Between) $>$ MS(Within) so $\frac{\text{MS(Between)}}{\text{MS(Within)}} > 1$

Testing in 1-way ANOVA (3 groups)

- $H_0 : \mu_1 = \mu_2 = \mu_3$, H_A : at least 2 means are not equal.

Testing in 1-way ANOVA (3 groups)

- $H_0 : \mu_1 = \mu_2 = \mu_3$, H_A : at least 2 means are not equal.

ANOVA table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Between	2	2.06337	1.03169	4.1047	0.04383
Error	12	3.01612	0.25134		

Testing in 1-way ANOVA (3 groups)

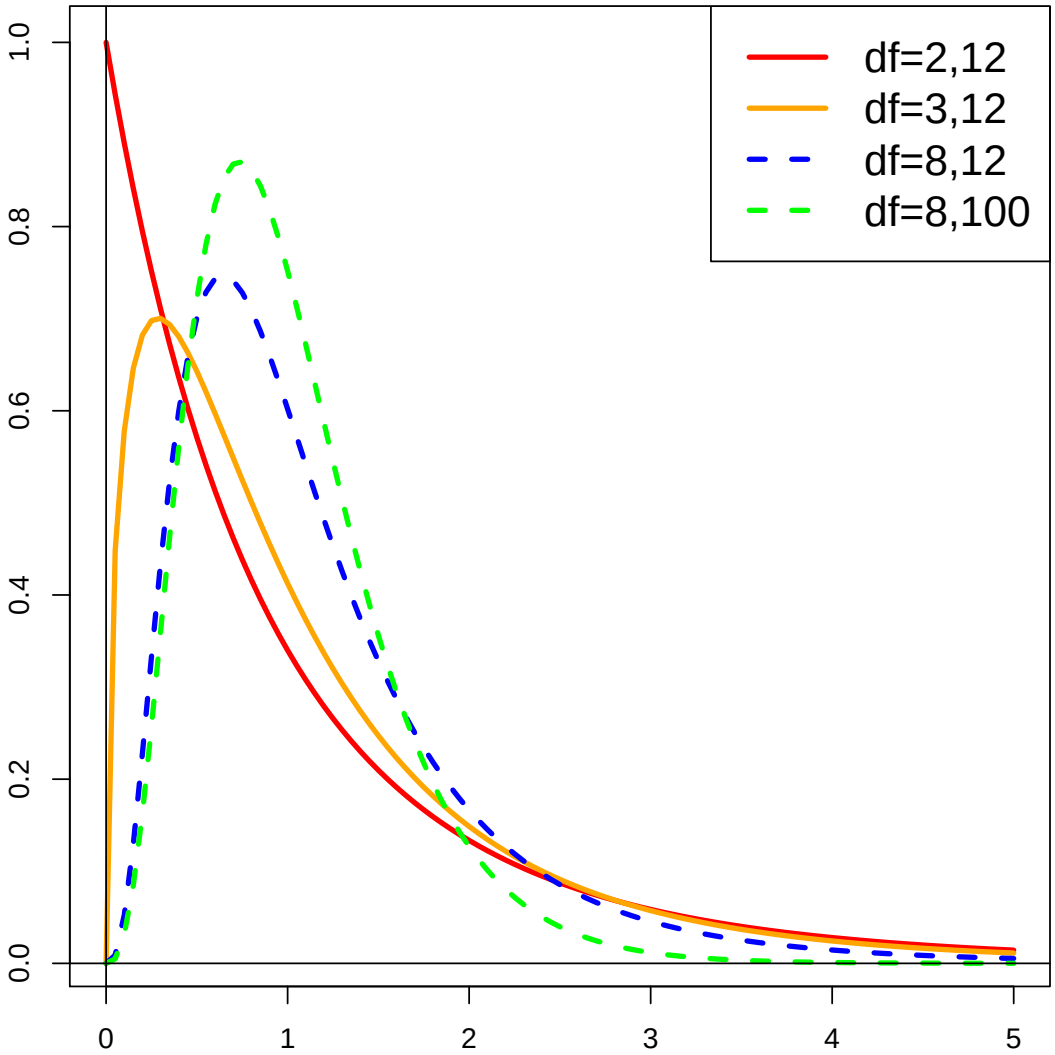
- $H_0 : \mu_1 = \mu_2 = \mu_3$, H_A : at least 2 means are not equal.

ANOVA table

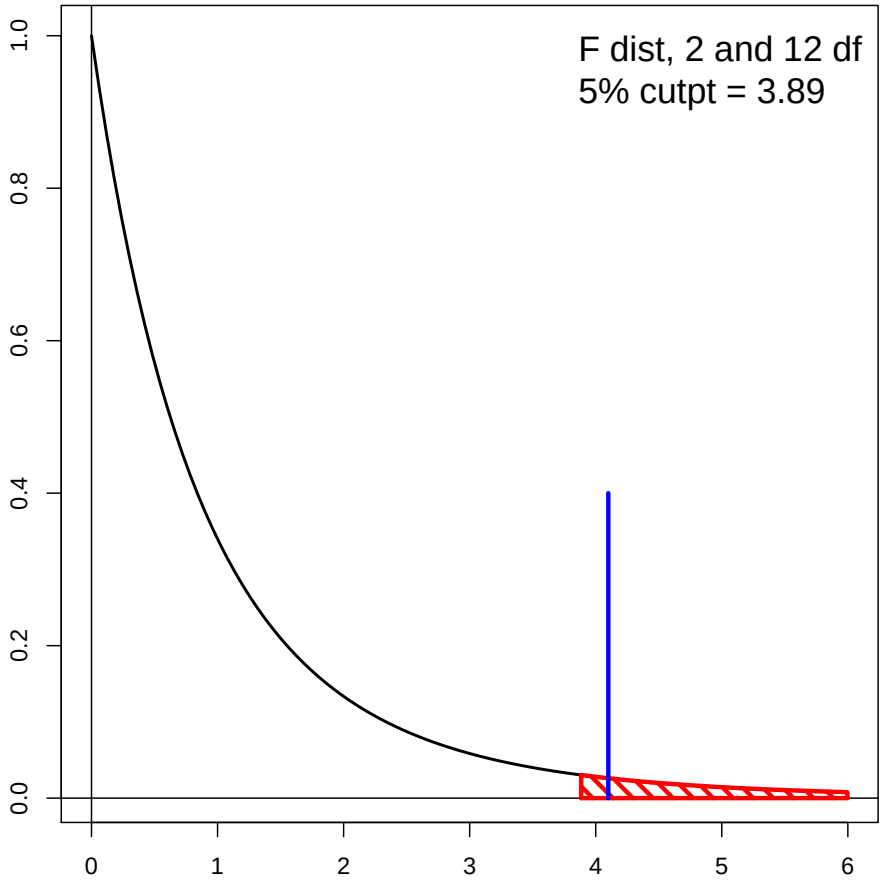
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Between	2	2.06337	1.03169	4.1047	0.04383
Error	12	3.01612	0.25134		

- Test of H_0 is $F_{calc} = \frac{\text{MS(Between)}}{\text{MS(Error)}} = 4.10$.
- If H_0 true (**and** assumptions met!), $F_{calc} \sim F_{2,12}$

Some F distributions



Dr. Crescendo's F_{calc}



- Generally: only large values of F_{calc} considered unusual (1-tail)
- Reject H_0 if $F_{calc} > F_{2,12,.95} = 3.89$.
- Dr. Crescendo: $F_{calc} = 1.03/0.25 = 4.10$, $p = 0.044$. Reject H_0.

Outline

Interacting with statistician

"What questions should I ask the statistician?"

Interacting with statistician

"What questions should I ask the statistician?"

- Who asks the first questions?
  - I ask: "Tell me about your experiment and your question."
  - We need some background
- If designing an experiment:
  - I ask: "What is your research question (explain in terms of data you plan to collect)?"
  - We can help you
    - Decide what data to collect to answer question
    - How to design experiment
    - Figure out reasonable sample size (but need further information from you: estimated effect)
- For data already collected
  - I ask: "What is your research question (explained in terms of data)?"
  - Plots of data helpful

How a statistician can help

- Design of experiment
  - Statistician must understand study aims and background
  - Can help investigator decide
    - what data to collect
    - how to design data collection process
    - how to use data to answer research questions
- Analysis
  - Determine appropriate model
  - Understand and check assumptions (use different model if violated)
  - Carry out analysis

What does Biostatistics Dept offer?

- Courses
  - BST463 = Introduction to Biostatistics (fall semester)
- Consulting service
  - Rotation system: faculty and PhD student or postdoc
  - Free assistance with grant preparation **if** appropriate % effort for Biostatistics
  - Free 1-hour consultation on anything
  - If related to CTSI, voucher possible for up to 10 free hours
  - Student projects: faculty mentor must attend initial meeting and most other meetings with Biostatisticians
  - Faculty advisor can contact Susan Messing (Susan_Messing@urmc.rochester.edu) for appointment, see http://www.urmc.rochester.edu/biostat/consulting/

Parting remarks

- Statistics is not just “crunching numbers”
  - Requires understanding of model and assumptions.
- Good experimental design and data collection is key
  - Poor design: may be impossible to make valid inference about population.
  - Biased data collection: may never be able to adjust for bias.
  - Good design: enables statistical inference about population of inference based on data collected.