



T Confidence Intervals

Statistical Inference

Brian Caffo, Jeff Leek, Roger Peng
Johns Hopkins Bloomberg School of Public Health

T Confidence intervals

- In the previous, we discussed creating a confidence interval using the CLT
 - They took the form $Est \pm ZQ \times SE_{Est}$
- In this lecture, we discuss some methods for small samples, notably Gosset's t distribution and t confidence intervals
 - They are of the form $Est \pm TQ \times SE_{Est}$
- These are some of the handiest of intervals
- If you want a rule between whether to use a t interval or normal interval, just always use the t interval
- We'll cover the one and two group versions

Gosset's t distribution

- Invented by William Gosset (under the pseudonym "Student") in 1908
- Has thicker tails than the normal
- Is indexed by degrees of freedom; gets more like a standard normal as df gets larger
- It assumes that the underlying data are iid Gaussian with the result that

$$\frac{\bar{X} - \mu}{S/\sqrt{n}}$$

follows Gosset's t distribution with $n - 1$ degrees of freedom

- (If we replaced s by σ the statistic would be exactly standard normal)
- Interval is $\bar{X} \pm t_{n-1} S/\sqrt{n}$ where t_{n-1} is the relevant quantile

Code for manipulate

```
k <- 1000
xvals <- seq(-5, 5, length = k)
myplot <- function(df){
  d <- data.frame(y = c(dnorm(xvals), dt(xvals, df)),
                  x = xvals,
                  dist = factor(rep(c("Normal", "T"), c(k,k))))
  g <- ggplot(d, aes(x = x, y = y))
  g <- g + geom_line(size = 2, aes(colour = dist))
  g
}
manipulate(myplot(mu), mu = slider(1, 20, step = 1))
```

Easier to see

```
pvals <- seq(.5, .99, by = .01)
myplot2 <- function(df){
  d <- data.frame(n= qnorm(pvals),t=qt(pvals, df),
                  p = pvals)
  g <- ggplot(d, aes(x= n, y = t))
  g <- g + geom_abline(size = 2, col = "lightblue")
  g <- g + geom_line(size = 2, col = "black")
  g <- g + geom_vline(xintercept = qnorm(0.975))
  g <- g + geom_hline(yintercept = qt(0.975, df))
  g
}
manipulate(myplot2(df), df = slider(1, 20, step = 1))
```

Note's about the t interval

- The t interval technically assumes that the data are iid normal, though it is robust to this assumption
- It works well whenever the distribution of the data is roughly symmetric and mound shaped
- Paired observations are often analyzed using the t interval by taking differences
- For large degrees of freedom, t quantiles become the same as standard normal quantiles; therefore this interval converges to the same interval as the CLT yielded
- For skewed distributions, the spirit of the t interval assumptions are violated
 - Also, for skewed distributions, it doesn't make a lot of sense to center the interval at the mean
 - In this case, consider taking logs or using a different summary like the median
- For highly discrete data, like binary, other intervals are available

Sleep data

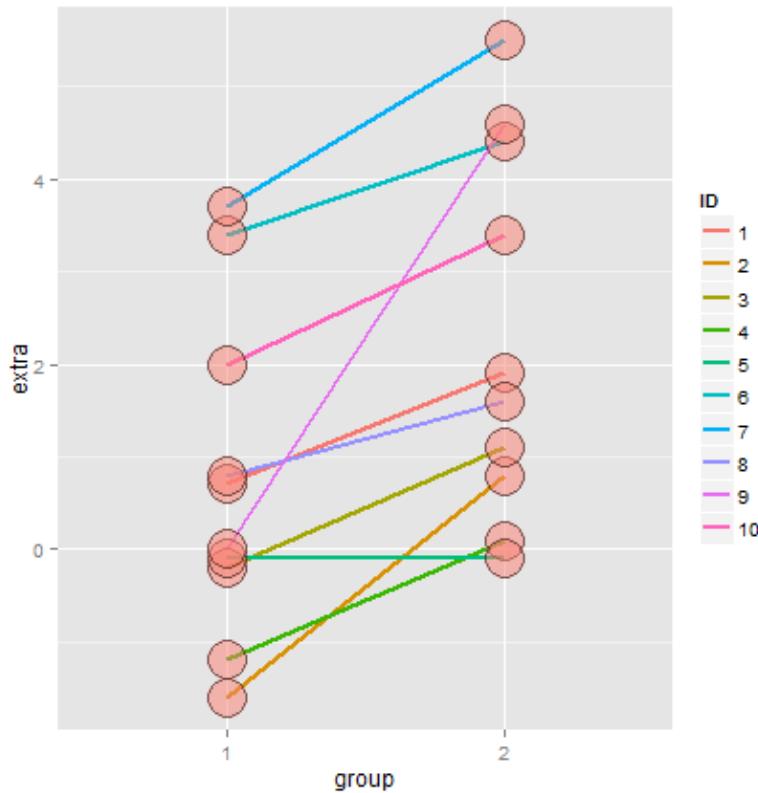
In R typing `data(sleep)` brings up the sleep data originally analyzed in Gosset's Biometrika paper, which shows the increase in hours for 10 patients on two soporific drugs. R treats the data as two groups rather than paired.

The data

```
data(sleep)
head(sleep)
```

```
##   extra group ID
## 1   0.7    1  1
## 2  -1.6    1  2
## 3  -0.2    1  3
## 4  -1.2    1  4
## 5  -0.1    1  5
## 6   3.4    1  6
```

Plotting the data



Results

```
g1 <- sleep$extra[1 : 10]; g2 <- sleep$extra[11 : 20]
difference <- g2 - g1
mn <- mean(difference); s <- sd(difference); n <- 10
```

```
mn + c(-1, 1) * qt(.975, n-1) * s / sqrt(n)
t.test(difference)
t.test(g2, g1, paired = TRUE)
t.test(extra ~ I(relevel(group, 2)), paired = TRUE, data = sleep)
```

The results

(After a little formatting)

```
##      [,1] [,2]
## [1,] 0.7001 2.46
## [2,] 0.7001 2.46
## [3,] 0.7001 2.46
## [4,] 0.7001 2.46
```

Independent group t confidence intervals

- Suppose that we want to compare the mean blood pressure between two groups in a randomized trial; those who received the treatment to those who received a placebo
- We cannot use the paired t test because the groups are independent and may have different sample sizes
- We now present methods for comparing independent groups

Confidence interval

- Therefore a $(1 - \alpha) \times 100\%$ confidence interval for $\mu_y - \mu_x$ is

$$\bar{Y} - \bar{X} \pm t_{n_x+n_y-2, 1-\alpha/2} S_p \left(\frac{1}{n_x} + \frac{1}{n_y} \right)^{1/2}$$

- The pooled variance estimator is

$$S_p^2 = \{(n_x - 1)S_x^2 + (n_y - 1)S_y^2\}/(n_x + n_y - 2)$$

- Remember this interval is assuming a constant variance across the two groups
- If there is some doubt, assume a different variance per group, which we will discuss later

Example

Based on Rosner, Fundamentals of Biostatistics

(Really a very good reference book)

- Comparing SBP for 8 oral contraceptive users versus 21 controls
- $\bar{X}_{OC} = 132.86$ mmHg with $s_{OC} = 15.34$ mmHg
- $\bar{X}_C = 127.44$ mmHg with $s_C = 18.23$ mmHg
- Pooled variance estimate

```
sp <- sqrt((7 * 15.34^2 + 20 * 18.23^2) / (8 + 21 - 2))
132.86 - 127.44 + c(-1, 1) * qt(.975, 27) * sp * (1 / 8 + 1 / 21)^.5
```

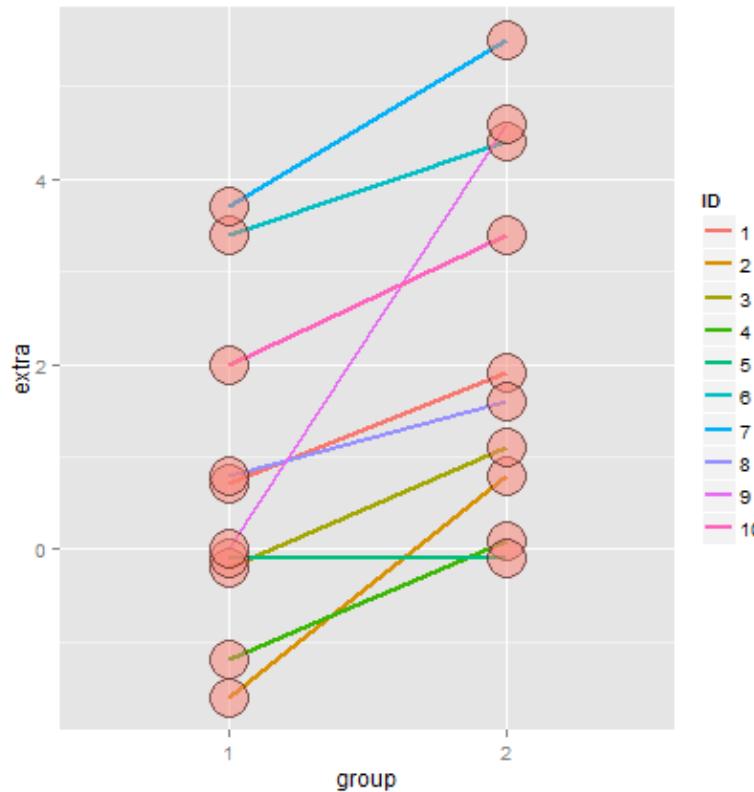
```
## [1] -9.521 20.361
```

Mistakenly treating the sleep data as grouped

```
n1 <- length(g1); n2 <- length(g2)
sp <- sqrt( ((n1 - 1) * sd(x1)^2 + (n2-1) * sd(x2)^2) / (n1 + n2-2) )
md <- mean(g2) - mean(g1)
semd <- sp * sqrt(1 / n1 + 1/n2)
rbind(
  md + c(-1, 1) * qt(.975, n1 + n2 - 2) * semd,
  t.test(g2, g1, paired = FALSE, var.equal = TRUE)$conf,
  t.test(g2, g1, paired = TRUE)$conf
)
```

```
##          [,1]  [,2]
## [1,] -0.2039 3.364
## [2,] -0.2039 3.364
## [3,]  0.7001 2.460
```

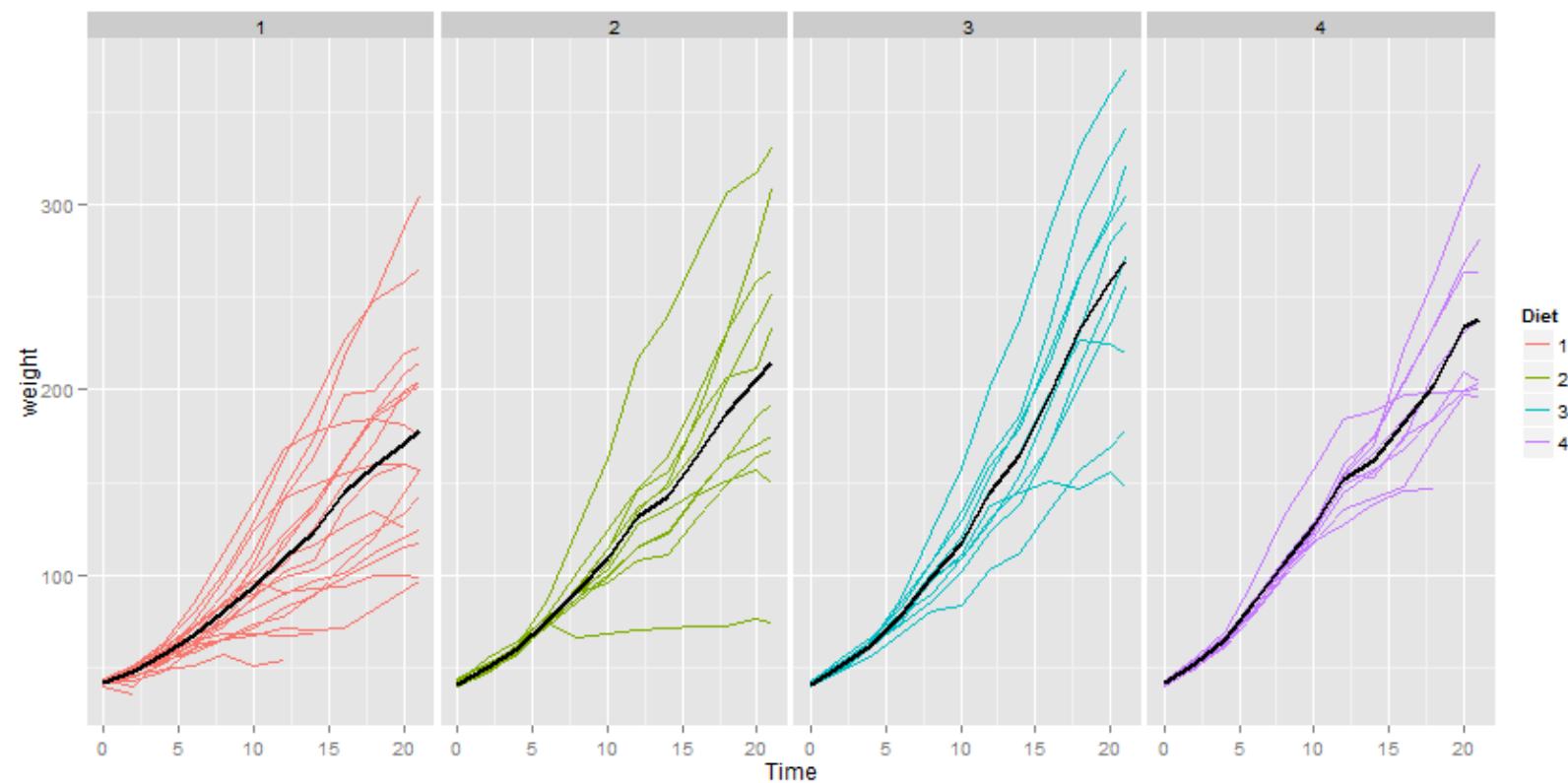
Grouped versus independent



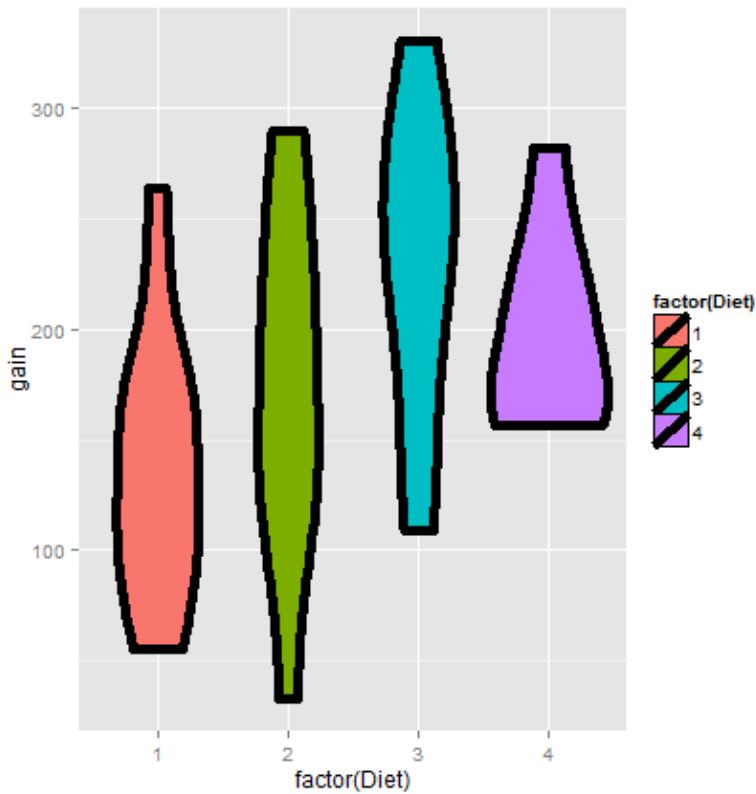
ChickWeight data in R

```
library(datasets); data(ChickWeight); library(reshape2)
##define weight gain or loss
wideCW <- dcast(ChickWeight, Diet + Chick ~ Time, value.var = "weight")
names(wideCW)[-c(1 : 2)] <- paste("time", names(wideCW)[-c(1 : 2)], sep = " ")
library(dplyr)
wideCW <- mutate(wideCW,
  gain = time21 - time0
)
```

Plotting the raw data



Weight gain by diet



Let's do a t interval

```
wideCW14 <- subset(wideCW, Diet %in% c(1, 4))
rbind(
t.test(gain ~ Diet, paired = FALSE, var.equal = TRUE, data = wideCW14)$conf,
t.test(gain ~ Diet, paired = FALSE, var.equal = FALSE, data = wideCW14)$conf
)
```

```
##          [,1]    [,2]
## [1,] -108.1 -14.81
## [2,] -104.7 -18.30
```

Unequal variances

- Under unequal variances

$$\bar{Y} - \bar{X} \pm t_{df} \times \left(\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y} \right)^{1/2}$$

where t_{df} is calculated with degrees of freedom

$$df = \frac{\left(S_x^2/n_x + S_y^2/n_y \right)^2}{\left(\frac{S_x^2}{n_x} \right)^2 / (n_x - 1) + \left(\frac{S_y^2}{n_y} \right)^2 / (n_y - 1)}$$

will be approximately a 95% interval

- This works really well
 - So when in doubt, just assume unequal variances

Example

- Comparing SBP for 8 oral contraceptive users versus 21 controls
- $\bar{X}_{OC} = 132.86 \text{ mmHg}$ with $s_{OC} = 15.34 \text{ mmHg}$
- $\bar{X}_C = 127.44 \text{ mmHg}$ with $s_C = 18.23 \text{ mmHg}$
- $df = 15.04, t_{15.04,.975} = 2.13$
- Interval

$$132.86 - 127.44 \pm 2.13 \left(\frac{15.34^2}{8} + \frac{18.23^2}{21} \right)^{1/2} = [-8.91, 19.75]$$

- In R, `t.test(..., var.equal = FALSE)`

Comparing other kinds of data

- For binomial data, there's lots of ways to compare two groups
 - Relative risk, risk difference, odds ratio.
 - Chi-squared tests, normal approximations, exact tests.
- For count data, there's also Chi-squared tests and exact tests.
- We'll leave the discussions for comparing groups of data for binary and count data until covering glms in the regression class.
- In addition, Mathematical Biostatistics Boot Camp 2 covers many special cases relevant to biostatistics.



T Confidence Intervals

Statistical Inference

Brian Caffo, Jeff Leek, Roger Peng
Johns Hopkins Bloomberg School of Public Health

T Confidence intervals

- In the previous, we discussed creating a confidence interval using the CLT
 - They took the form $Est \pm ZQ \times SE_{Est}$
- In this lecture, we discuss some methods for small samples, notably Gosset's t distribution and t confidence intervals
 - They are of the form $Est \pm TQ \times SE_{Est}$
- These are some of the handiest of intervals
- If you want a rule between whether to use a t interval or normal interval, just always use the t interval
- We'll cover the one and two group versions

Gosset's t distribution

- Invented by William Gosset (under the pseudonym "Student") in 1908
- Has thicker tails than the normal
- Is indexed by degrees of freedom; gets more like a standard normal as df gets larger
- It assumes that the underlying data are iid Gaussian with the result that

$$\frac{\bar{X} - \mu}{S/\sqrt{n}}$$

follows Gosset's t distribution with $n - 1$ degrees of freedom

- (If we replaced s by σ the statistic would be exactly standard normal)
- Interval is $\bar{X} \pm t_{n-1} S/\sqrt{n}$ where t_{n-1} is the relevant quantile

Code for manipulate

```
k <- 1000
xvals <- seq(-5, 5, length = k)
myplot <- function(df){
  d <- data.frame(y = c(dnorm(xvals), dt(xvals, df)),
                  x = xvals,
                  dist = factor(rep(c("Normal", "T"), c(k,k))))
  g <- ggplot(d, aes(x = x, y = y))
  g <- g + geom_line(size = 2, aes(colour = dist))
  g
}
manipulate(myplot(mu), mu = slider(1, 20, step = 1))
```

Easier to see

```
pvals <- seq(.5, .99, by = .01)
myplot2 <- function(df){
  d <- data.frame(n= qnorm(pvals),t=qt(pvals, df),
                  p = pvals)
  g <- ggplot(d, aes(x= n, y = t))
  g <- g + geom_abline(size = 2, col = "lightblue")
  g <- g + geom_line(size = 2, col = "black")
  g <- g + geom_vline(xintercept = qnorm(0.975))
  g <- g + geom_hline(yintercept = qt(0.975, df))
  g
}
manipulate(myplot2(df), df = slider(1, 20, step = 1))
```

Note's about the t interval

- The t interval technically assumes that the data are iid normal, though it is robust to this assumption
- It works well whenever the distribution of the data is roughly symmetric and mound shaped
- Paired observations are often analyzed using the t interval by taking differences
- For large degrees of freedom, t quantiles become the same as standard normal quantiles; therefore this interval converges to the same interval as the CLT yielded
- For skewed distributions, the spirit of the t interval assumptions are violated
 - Also, for skewed distributions, it doesn't make a lot of sense to center the interval at the mean
 - In this case, consider taking logs or using a different summary like the median
- For highly discrete data, like binary, other intervals are available

Sleep data

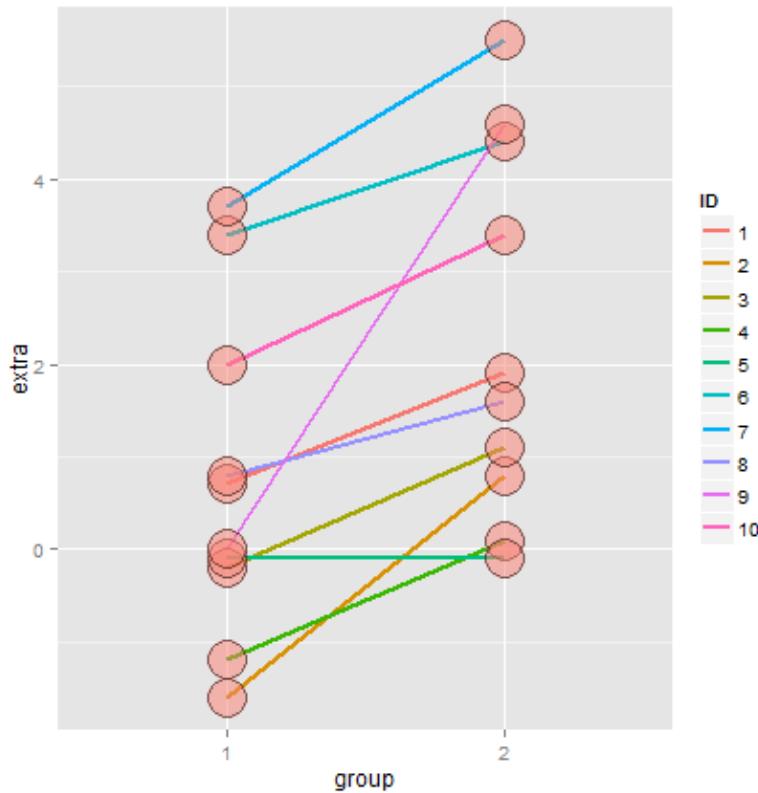
In R typing `data(sleep)` brings up the sleep data originally analyzed in Gosset's Biometrika paper, which shows the increase in hours for 10 patients on two soporific drugs. R treats the data as two groups rather than paired.

The data

```
data(sleep)
head(sleep)
```

```
##   extra group ID
## 1   0.7    1  1
## 2  -1.6    1  2
## 3  -0.2    1  3
## 4  -1.2    1  4
## 5  -0.1    1  5
## 6   3.4    1  6
```

Plotting the data



Results

```
g1 <- sleep$extra[1 : 10]; g2 <- sleep$extra[11 : 20]
difference <- g2 - g1
mn <- mean(difference); s <- sd(difference); n <- 10
```

```
mn + c(-1, 1) * qt(.975, n-1) * s / sqrt(n)
t.test(difference)
t.test(g2, g1, paired = TRUE)
t.test(extra ~ I(relevel(group, 2)), paired = TRUE, data = sleep)
```

The results

(After a little formatting)

```
##      [,1] [,2]
## [1,] 0.7001 2.46
## [2,] 0.7001 2.46
## [3,] 0.7001 2.46
## [4,] 0.7001 2.46
```

Independent group t confidence intervals

- Suppose that we want to compare the mean blood pressure between two groups in a randomized trial; those who received the treatment to those who received a placebo
- We cannot use the paired t test because the groups are independent and may have different sample sizes
- We now present methods for comparing independent groups

Confidence interval

- Therefore a $(1 - \alpha) \times 100\%$ confidence interval for $\mu_y - \mu_x$ is

$$\bar{Y} - \bar{X} \pm t_{n_x+n_y-2, 1-\alpha/2} S_p \left(\frac{1}{n_x} + \frac{1}{n_y} \right)^{1/2}$$

- The pooled variance estimator is

$$S_p^2 = \{(n_x - 1)S_x^2 + (n_y - 1)S_y^2\}/(n_x + n_y - 2)$$

- Remember this interval is assuming a constant variance across the two groups
- If there is some doubt, assume a different variance per group, which we will discuss later

Example

Based on Rosner, Fundamentals of Biostatistics

(Really a very good reference book)

- Comparing SBP for 8 oral contraceptive users versus 21 controls
- $\bar{X}_{OC} = 132.86$ mmHg with $s_{OC} = 15.34$ mmHg
- $\bar{X}_C = 127.44$ mmHg with $s_C = 18.23$ mmHg
- Pooled variance estimate

```
sp <- sqrt((7 * 15.34^2 + 20 * 18.23^2) / (8 + 21 - 2))
132.86 - 127.44 + c(-1, 1) * qt(.975, 27) * sp * (1 / 8 + 1 / 21)^.5
```

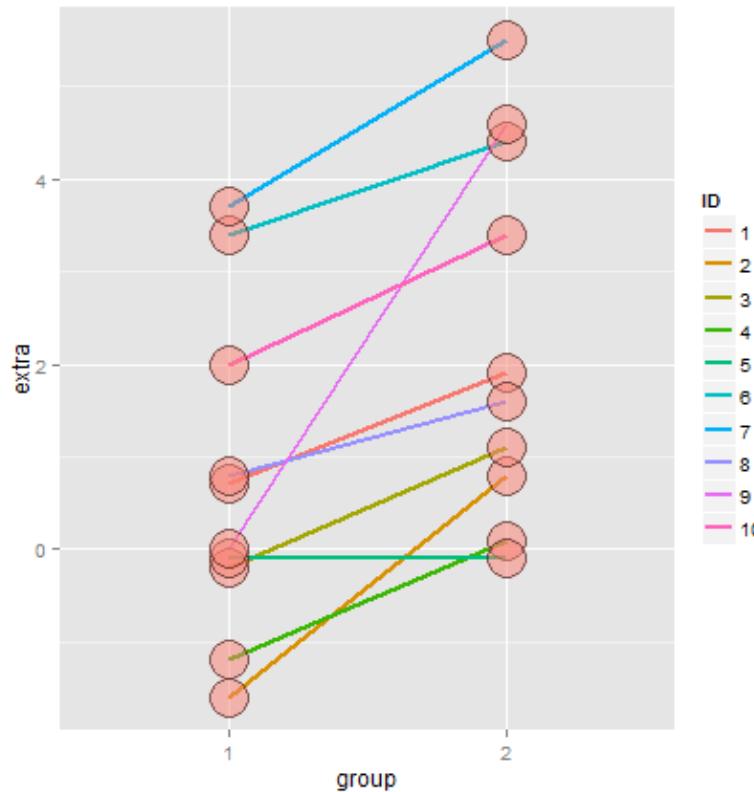
```
## [1] -9.521 20.361
```

Mistakenly treating the sleep data as grouped

```
n1 <- length(g1); n2 <- length(g2)
sp <- sqrt( ((n1 - 1) * sd(x1)^2 + (n2-1) * sd(x2)^2) / (n1 + n2-2) )
md <- mean(g2) - mean(g1)
semd <- sp * sqrt(1 / n1 + 1/n2)
rbind(
  md + c(-1, 1) * qt(.975, n1 + n2 - 2) * semd,
  t.test(g2, g1, paired = FALSE, var.equal = TRUE)$conf,
  t.test(g2, g1, paired = TRUE)$conf
)
```

```
##          [,1]  [,2]
## [1,] -0.2039 3.364
## [2,] -0.2039 3.364
## [3,]  0.7001 2.460
```

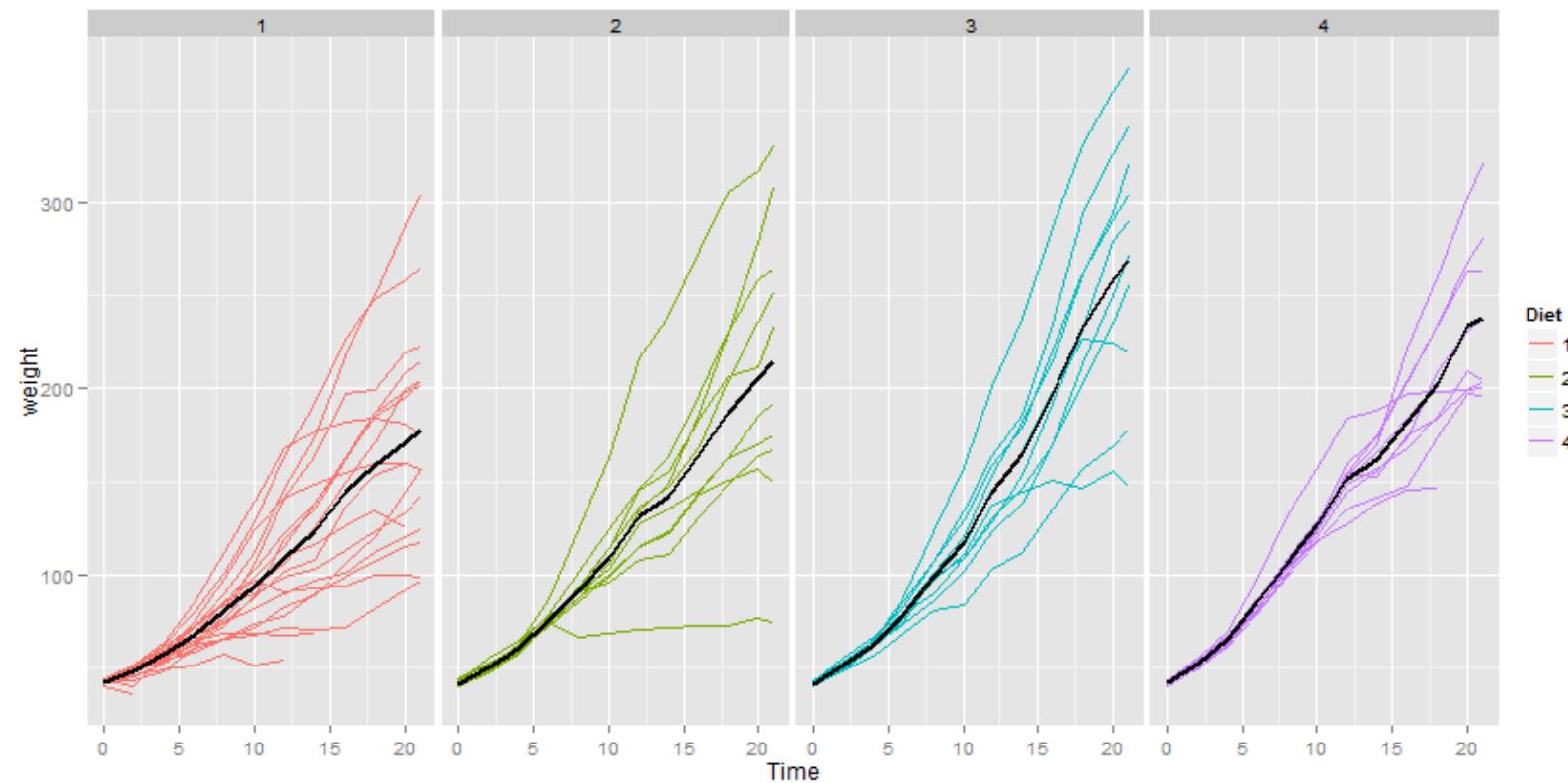
Grouped versus independent



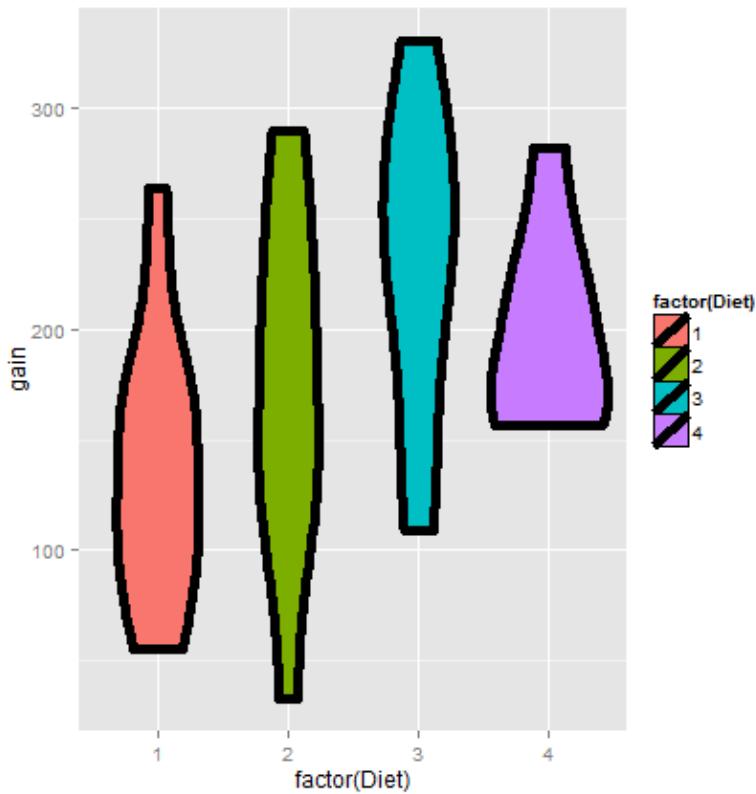
ChickWeight data in R

```
library(datasets); data(ChickWeight); library(reshape2)
##define weight gain or loss
wideCW <- dcast(ChickWeight, Diet + Chick ~ Time, value.var = "weight")
names(wideCW)[-c(1 : 2)] <- paste("time", names(wideCW)[-c(1 : 2)], sep = " ")
library(dplyr)
wideCW <- mutate(wideCW,
  gain = time21 - time0
)
```

Plotting the raw data



Weight gain by diet



Let's do a t interval

```
wideCW14 <- subset(wideCW, Diet %in% c(1, 4))
rbind(
t.test(gain ~ Diet, paired = FALSE, var.equal = TRUE, data = wideCW14)$conf,
t.test(gain ~ Diet, paired = FALSE, var.equal = FALSE, data = wideCW14)$conf
)
```

```
##          [,1]    [,2]
## [1,] -108.1 -14.81
## [2,] -104.7 -18.30
```

Unequal variances

- Under unequal variances

$$\bar{Y} - \bar{X} \pm t_{df} \times \left(\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y} \right)^{1/2}$$

where t_{df} is calculated with degrees of freedom

$$df = \frac{\left(S_x^2/n_x + S_y^2/n_y \right)^2}{\left(\frac{S_x^2}{n_x} \right)^2 / (n_x - 1) + \left(\frac{S_y^2}{n_y} \right)^2 / (n_y - 1)}$$

will be approximately a 95% interval

- This works really well
 - So when in doubt, just assume unequal variances

Example

- Comparing SBP for 8 oral contraceptive users versus 21 controls
- $\bar{X}_{OC} = 132.86 \text{ mmHg}$ with $s_{OC} = 15.34 \text{ mmHg}$
- $\bar{X}_C = 127.44 \text{ mmHg}$ with $s_C = 18.23 \text{ mmHg}$
- $df = 15.04, t_{15.04,.975} = 2.13$
- Interval

$$132.86 - 127.44 \pm 2.13 \left(\frac{15.34^2}{8} + \frac{18.23^2}{21} \right)^{1/2} = [-8.91, 19.75]$$

- In R, `t.test(..., var.equal = FALSE)`

Comparing other kinds of data

- For binomial data, there's lots of ways to compare two groups
 - Relative risk, risk difference, odds ratio.
 - Chi-squared tests, normal approximations, exact tests.
- For count data, there's also Chi-squared tests and exact tests.
- We'll leave the discussions for comparing groups of data for binary and count data until covering glms in the regression class.
- In addition, Mathematical Biostatistics Boot Camp 2 covers many special cases relevant to biostatistics.



T Confidence Intervals

Statistical Inference

Brian Caffo, Jeff Leek, Roger Peng
Johns Hopkins Bloomberg School of Public Health

T Confidence intervals

- In the previous, we discussed creating a confidence interval using the CLT
 - They took the form $Est \pm ZQ \times SE_{Est}$
- In this lecture, we discuss some methods for small samples, notably Gosset's t distribution and t confidence intervals
 - They are of the form $Est \pm TQ \times SE_{Est}$
- These are some of the handiest of intervals
- If you want a rule between whether to use a t interval or normal interval, just always use the t interval
- We'll cover the one and two group versions

Gosset's t distribution

- Invented by William Gosset (under the pseudonym "Student") in 1908
- Has thicker tails than the normal
- Is indexed by degrees of freedom; gets more like a standard normal as df gets larger
- It assumes that the underlying data are iid Gaussian with the result that

$$\frac{\bar{X} - \mu}{S/\sqrt{n}}$$

follows Gosset's t distribution with $n - 1$ degrees of freedom

- (If we replaced s by σ the statistic would be exactly standard normal)
- Interval is $\bar{X} \pm t_{n-1} S/\sqrt{n}$ where t_{n-1} is the relevant quantile

Code for manipulate

```
k <- 1000
xvals <- seq(-5, 5, length = k)
myplot <- function(df){
  d <- data.frame(y = c(dnorm(xvals), dt(xvals, df)),
                  x = xvals,
                  dist = factor(rep(c("Normal", "T"), c(k,k))))
  g <- ggplot(d, aes(x = x, y = y))
  g <- g + geom_line(size = 2, aes(colour = dist))
  g
}
manipulate(myplot(mu), mu = slider(1, 20, step = 1))
```

Easier to see

```
pvals <- seq(.5, .99, by = .01)
myplot2 <- function(df){
  d <- data.frame(n= qnorm(pvals),t=qt(pvals, df),
                  p = pvals)
  g <- ggplot(d, aes(x= n, y = t))
  g <- g + geom_abline(size = 2, col = "lightblue")
  g <- g + geom_line(size = 2, col = "black")
  g <- g + geom_vline(xintercept = qnorm(0.975))
  g <- g + geom_hline(yintercept = qt(0.975, df))
  g
}
manipulate(myplot2(df), df = slider(1, 20, step = 1))
```

Note's about the t interval

- The t interval technically assumes that the data are iid normal, though it is robust to this assumption
- It works well whenever the distribution of the data is roughly symmetric and mound shaped
- Paired observations are often analyzed using the t interval by taking differences
- For large degrees of freedom, t quantiles become the same as standard normal quantiles; therefore this interval converges to the same interval as the CLT yielded
- For skewed distributions, the spirit of the t interval assumptions are violated
 - Also, for skewed distributions, it doesn't make a lot of sense to center the interval at the mean
 - In this case, consider taking logs or using a different summary like the median
- For highly discrete data, like binary, other intervals are available

Sleep data

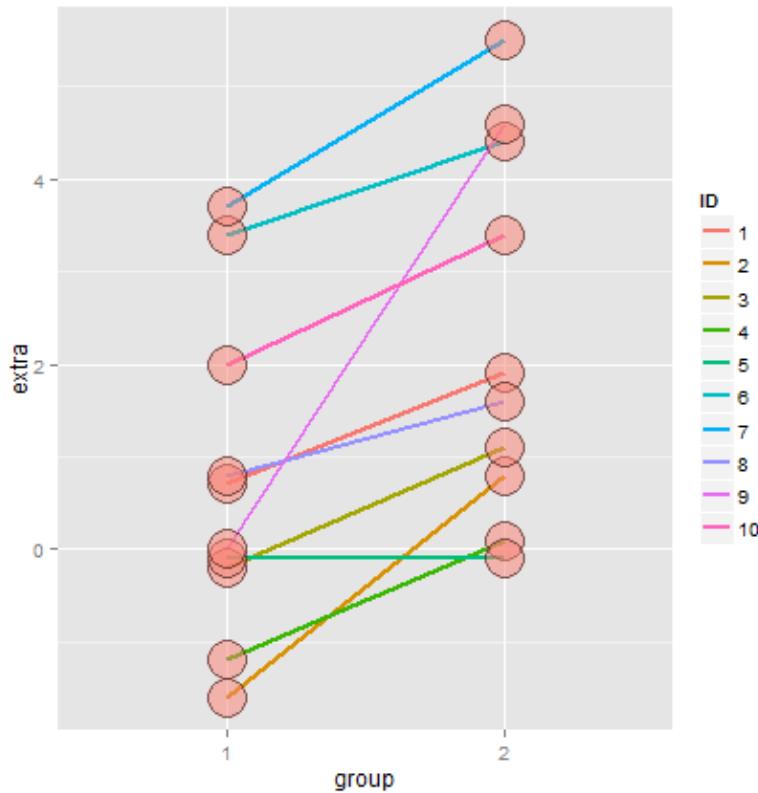
In R typing `data(sleep)` brings up the sleep data originally analyzed in Gosset's Biometrika paper, which shows the increase in hours for 10 patients on two soporific drugs. R treats the data as two groups rather than paired.

The data

```
data(sleep)
head(sleep)
```

```
##   extra group ID
## 1   0.7    1  1
## 2  -1.6    1  2
## 3  -0.2    1  3
## 4  -1.2    1  4
## 5  -0.1    1  5
## 6   3.4    1  6
```

Plotting the data



Results

```
g1 <- sleep$extra[1 : 10]; g2 <- sleep$extra[11 : 20]
difference <- g2 - g1
mn <- mean(difference); s <- sd(difference); n <- 10
```

```
mn + c(-1, 1) * qt(.975, n-1) * s / sqrt(n)
t.test(difference)
t.test(g2, g1, paired = TRUE)
t.test(extra ~ I(relevel(group, 2)), paired = TRUE, data = sleep)
```

The results

(After a little formatting)

```
##      [,1] [,2]
## [1,] 0.7001 2.46
## [2,] 0.7001 2.46
## [3,] 0.7001 2.46
## [4,] 0.7001 2.46
```

Independent group t confidence intervals

- Suppose that we want to compare the mean blood pressure between two groups in a randomized trial; those who received the treatment to those who received a placebo
- We cannot use the paired t test because the groups are independent and may have different sample sizes
- We now present methods for comparing independent groups

Confidence interval

- Therefore a $(1 - \alpha) \times 100\%$ confidence interval for $\mu_y - \mu_x$ is

$$\bar{Y} - \bar{X} \pm t_{n_x+n_y-2, 1-\alpha/2} S_p \left(\frac{1}{n_x} + \frac{1}{n_y} \right)^{1/2}$$

- The pooled variance estimator is

$$S_p^2 = \{(n_x - 1)S_x^2 + (n_y - 1)S_y^2\}/(n_x + n_y - 2)$$

- Remember this interval is assuming a constant variance across the two groups
- If there is some doubt, assume a different variance per group, which we will discuss later

Example

Based on Rosner, Fundamentals of Biostatistics

(Really a very good reference book)

- Comparing SBP for 8 oral contraceptive users versus 21 controls
- $\bar{X}_{OC} = 132.86$ mmHg with $s_{OC} = 15.34$ mmHg
- $\bar{X}_C = 127.44$ mmHg with $s_C = 18.23$ mmHg
- Pooled variance estimate

```
sp <- sqrt((7 * 15.34^2 + 20 * 18.23^2) / (8 + 21 - 2))
132.86 - 127.44 + c(-1, 1) * qt(.975, 27) * sp * (1 / 8 + 1 / 21)^.5
```

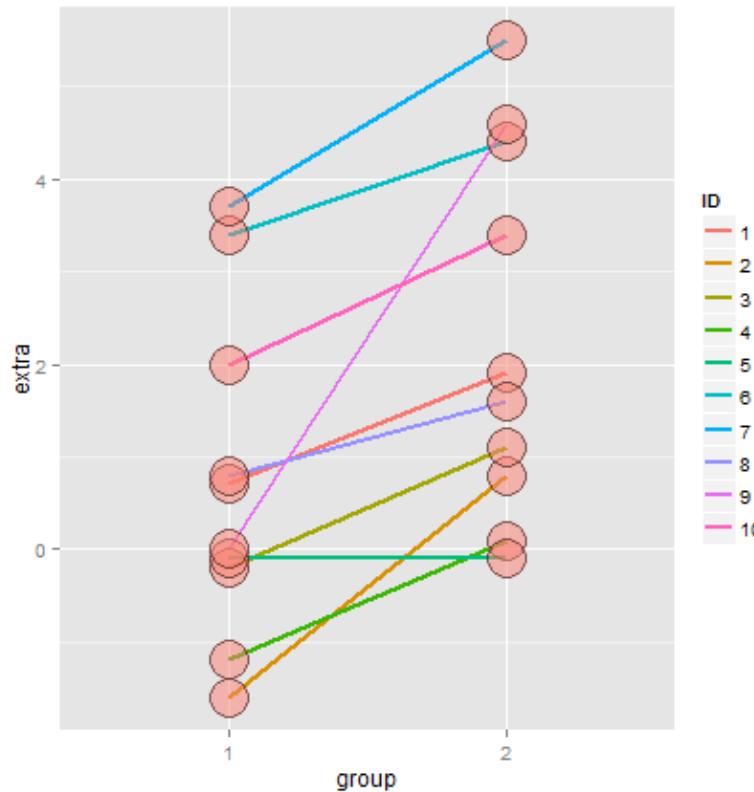
```
## [1] -9.521 20.361
```

Mistakenly treating the sleep data as grouped

```
n1 <- length(g1); n2 <- length(g2)
sp <- sqrt( ((n1 - 1) * sd(x1)^2 + (n2-1) * sd(x2)^2) / (n1 + n2-2) )
md <- mean(g2) - mean(g1)
semd <- sp * sqrt(1 / n1 + 1/n2)
rbind(
  md + c(-1, 1) * qt(.975, n1 + n2 - 2) * semd,
  t.test(g2, g1, paired = FALSE, var.equal = TRUE)$conf,
  t.test(g2, g1, paired = TRUE)$conf
)
```

```
##          [,1]  [,2]
## [1,] -0.2039 3.364
## [2,] -0.2039 3.364
## [3,]  0.7001 2.460
```

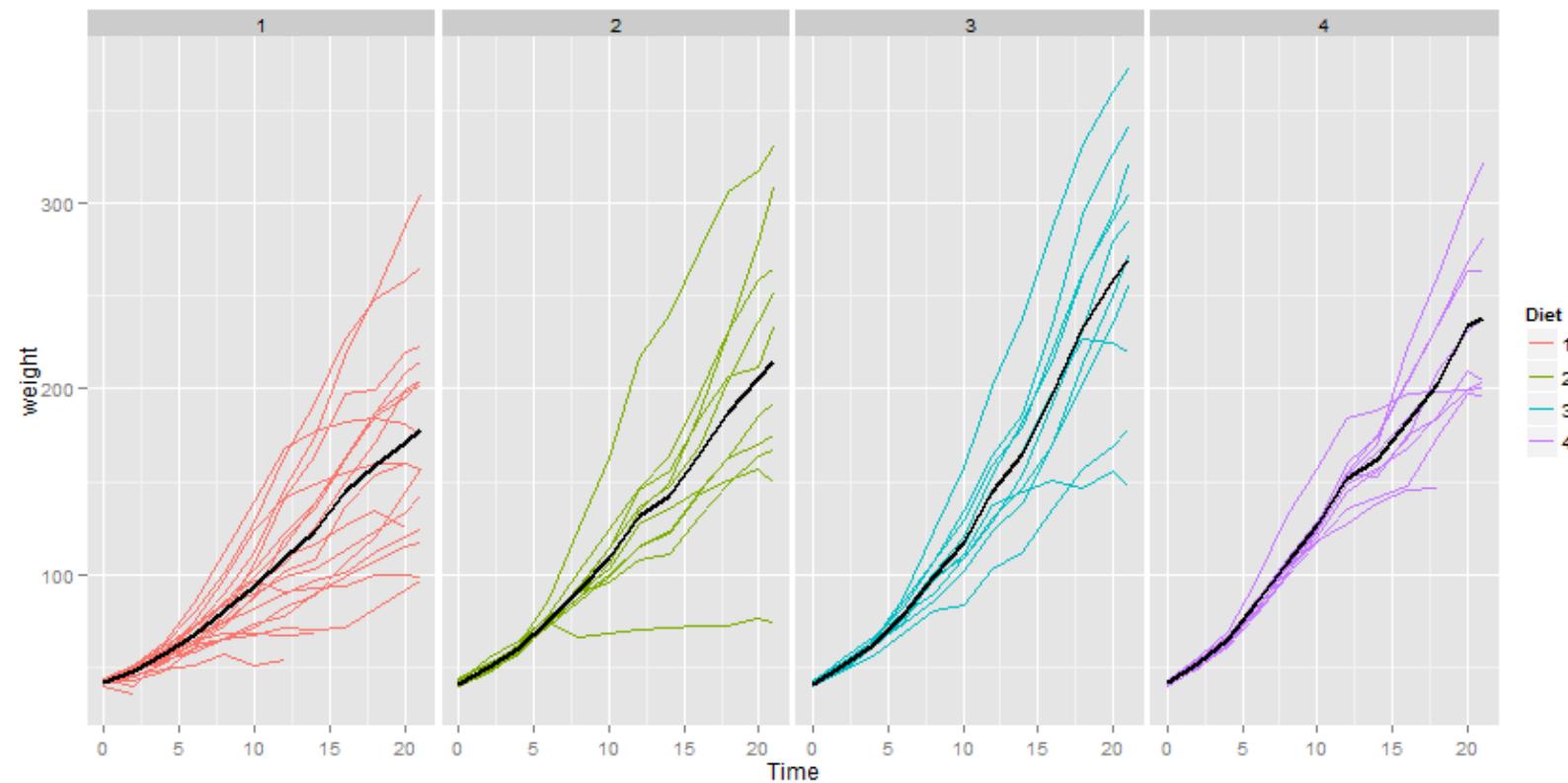
Grouped versus independent



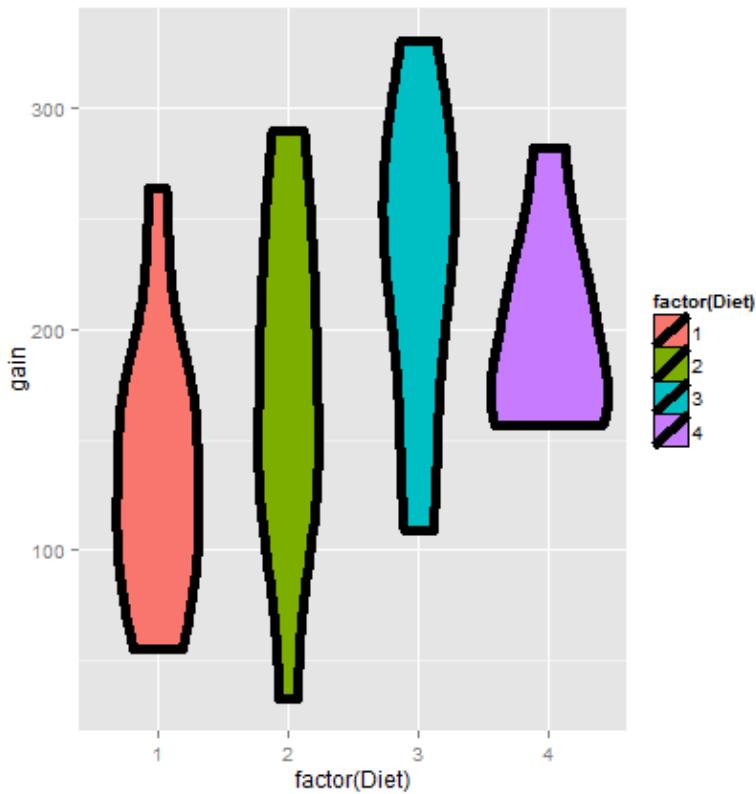
ChickWeight data in R

```
library(datasets); data(ChickWeight); library(reshape2)
##define weight gain or loss
wideCW <- dcast(ChickWeight, Diet + Chick ~ Time, value.var = "weight")
names(wideCW)[-c(1 : 2)] <- paste("time", names(wideCW)[-c(1 : 2)], sep = " ")
library(dplyr)
wideCW <- mutate(wideCW,
  gain = time21 - time0
)
```

Plotting the raw data



Weight gain by diet



Let's do a t interval

```
wideCW14 <- subset(wideCW, Diet %in% c(1, 4))
rbind(
t.test(gain ~ Diet, paired = FALSE, var.equal = TRUE, data = wideCW14)$conf,
t.test(gain ~ Diet, paired = FALSE, var.equal = FALSE, data = wideCW14)$conf
)
```

```
##          [,1]    [,2]
## [1,] -108.1 -14.81
## [2,] -104.7 -18.30
```

Unequal variances

- Under unequal variances

$$\bar{Y} - \bar{X} \pm t_{df} \times \left(\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y} \right)^{1/2}$$

where t_{df} is calculated with degrees of freedom

$$df = \frac{\left(S_x^2/n_x + S_y^2/n_y \right)^2}{\left(\frac{S_x^2}{n_x} \right)^2 / (n_x - 1) + \left(\frac{S_y^2}{n_y} \right)^2 / (n_y - 1)}$$

will be approximately a 95% interval

- This works really well
 - So when in doubt, just assume unequal variances

Example

- Comparing SBP for 8 oral contraceptive users versus 21 controls
- $\bar{X}_{OC} = 132.86 \text{ mmHg}$ with $s_{OC} = 15.34 \text{ mmHg}$
- $\bar{X}_C = 127.44 \text{ mmHg}$ with $s_C = 18.23 \text{ mmHg}$
- $df = 15.04, t_{15.04,.975} = 2.13$
- Interval

$$132.86 - 127.44 \pm 2.13 \left(\frac{15.34^2}{8} + \frac{18.23^2}{21} \right)^{1/2} = [-8.91, 19.75]$$

- In R, `t.test(..., var.equal = FALSE)`

Comparing other kinds of data

- For binomial data, there's lots of ways to compare two groups
 - Relative risk, risk difference, odds ratio.
 - Chi-squared tests, normal approximations, exact tests.
- For count data, there's also Chi-squared tests and exact tests.
- We'll leave the discussions for comparing groups of data for binary and count data until covering glms in the regression class.
- In addition, Mathematical Biostatistics Boot Camp 2 covers many special cases relevant to biostatistics.



T Confidence Intervals

Statistical Inference

Brian Caffo, Jeff Leek, Roger Peng
Johns Hopkins Bloomberg School of Public Health

T Confidence intervals

- In the previous, we discussed creating a confidence interval using the CLT
 - They took the form $Est \pm ZQ \times SE_{Est}$
- In this lecture, we discuss some methods for small samples, notably Gosset's t distribution and t confidence intervals
 - They are of the form $Est \pm TQ \times SE_{Est}$
- These are some of the handiest of intervals
- If you want a rule between whether to use a t interval or normal interval, just always use the t interval
- We'll cover the one and two group versions

Gosset's t distribution

- Invented by William Gosset (under the pseudonym "Student") in 1908
- Has thicker tails than the normal
- Is indexed by degrees of freedom; gets more like a standard normal as df gets larger
- It assumes that the underlying data are iid Gaussian with the result that

$$\frac{\bar{X} - \mu}{S/\sqrt{n}}$$

follows Gosset's t distribution with $n - 1$ degrees of freedom

- (If we replaced s by σ the statistic would be exactly standard normal)
- Interval is $\bar{X} \pm t_{n-1} S/\sqrt{n}$ where t_{n-1} is the relevant quantile

Code for manipulate

```
k <- 1000
xvals <- seq(-5, 5, length = k)
myplot <- function(df){
  d <- data.frame(y = c(dnorm(xvals), dt(xvals, df)),
                  x = xvals,
                  dist = factor(rep(c("Normal", "T"), c(k,k))))
  g <- ggplot(d, aes(x = x, y = y))
  g <- g + geom_line(size = 2, aes(colour = dist))
  g
}
manipulate(myplot(mu), mu = slider(1, 20, step = 1))
```

Easier to see

```
pvals <- seq(.5, .99, by = .01)
myplot2 <- function(df){
  d <- data.frame(n= qnorm(pvals),t=qt(pvals, df),
                  p = pvals)
  g <- ggplot(d, aes(x= n, y = t))
  g <- g + geom_abline(size = 2, col = "lightblue")
  g <- g + geom_line(size = 2, col = "black")
  g <- g + geom_vline(xintercept = qnorm(0.975))
  g <- g + geom_hline(yintercept = qt(0.975, df))
  g
}
manipulate(myplot2(df), df = slider(1, 20, step = 1))
```

Note's about the t interval

- The t interval technically assumes that the data are iid normal, though it is robust to this assumption
- It works well whenever the distribution of the data is roughly symmetric and mound shaped
- Paired observations are often analyzed using the t interval by taking differences
- For large degrees of freedom, t quantiles become the same as standard normal quantiles; therefore this interval converges to the same interval as the CLT yielded
- For skewed distributions, the spirit of the t interval assumptions are violated
 - Also, for skewed distributions, it doesn't make a lot of sense to center the interval at the mean
 - In this case, consider taking logs or using a different summary like the median
- For highly discrete data, like binary, other intervals are available

Sleep data

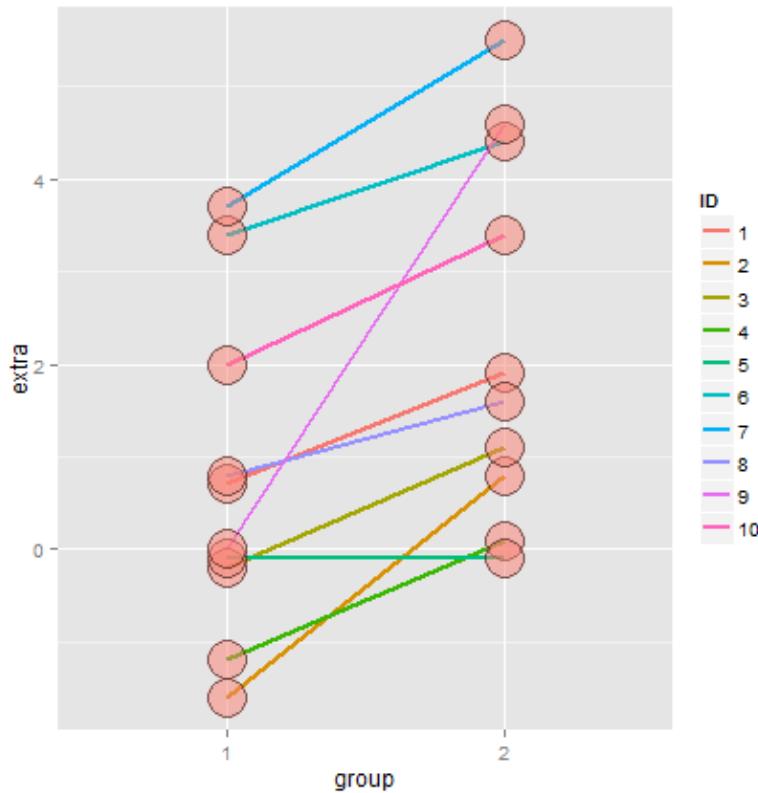
In R typing `data(sleep)` brings up the sleep data originally analyzed in Gosset's Biometrika paper, which shows the increase in hours for 10 patients on two soporific drugs. R treats the data as two groups rather than paired.

The data

```
data(sleep)
head(sleep)
```

```
##   extra group ID
## 1   0.7    1  1
## 2  -1.6    1  2
## 3  -0.2    1  3
## 4  -1.2    1  4
## 5  -0.1    1  5
## 6   3.4    1  6
```

Plotting the data



Results

```
g1 <- sleep$extra[1 : 10]; g2 <- sleep$extra[11 : 20]
difference <- g2 - g1
mn <- mean(difference); s <- sd(difference); n <- 10
```

```
mn + c(-1, 1) * qt(.975, n-1) * s / sqrt(n)
t.test(difference)
t.test(g2, g1, paired = TRUE)
t.test(extra ~ I(relevel(group, 2)), paired = TRUE, data = sleep)
```

The results

(After a little formatting)

```
##      [,1] [,2]
## [1,] 0.7001 2.46
## [2,] 0.7001 2.46
## [3,] 0.7001 2.46
## [4,] 0.7001 2.46
```

Independent group t confidence intervals

- Suppose that we want to compare the mean blood pressure between two groups in a randomized trial; those who received the treatment to those who received a placebo
- We cannot use the paired t test because the groups are independent and may have different sample sizes
- We now present methods for comparing independent groups

Confidence interval

- Therefore a $(1 - \alpha) \times 100\%$ confidence interval for $\mu_y - \mu_x$ is

$$\bar{Y} - \bar{X} \pm t_{n_x+n_y-2, 1-\alpha/2} S_p \left(\frac{1}{n_x} + \frac{1}{n_y} \right)^{1/2}$$

- The pooled variance estimator is

$$S_p^2 = \{(n_x - 1)S_x^2 + (n_y - 1)S_y^2\}/(n_x + n_y - 2)$$

- Remember this interval is assuming a constant variance across the two groups
- If there is some doubt, assume a different variance per group, which we will discuss later

Example

Based on Rosner, Fundamentals of Biostatistics

(Really a very good reference book)

- Comparing SBP for 8 oral contraceptive users versus 21 controls
- $\bar{X}_{OC} = 132.86$ mmHg with $s_{OC} = 15.34$ mmHg
- $\bar{X}_C = 127.44$ mmHg with $s_C = 18.23$ mmHg
- Pooled variance estimate

```
sp <- sqrt((7 * 15.34^2 + 20 * 18.23^2) / (8 + 21 - 2))
132.86 - 127.44 + c(-1, 1) * qt(.975, 27) * sp * (1 / 8 + 1 / 21)^.5
```

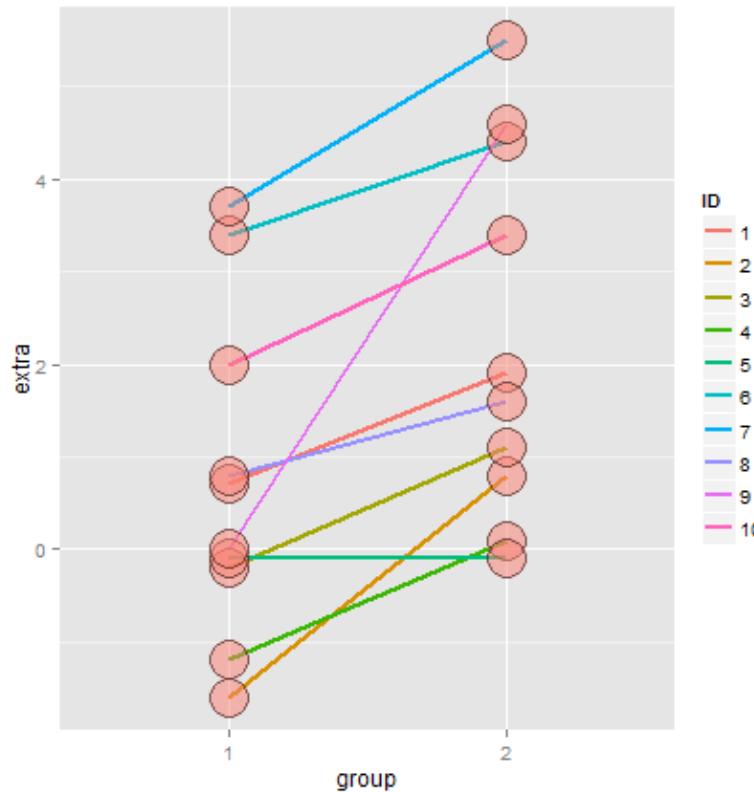
```
## [1] -9.521 20.361
```

Mistakenly treating the sleep data as grouped

```
n1 <- length(g1); n2 <- length(g2)
sp <- sqrt( ((n1 - 1) * sd(x1)^2 + (n2-1) * sd(x2)^2) / (n1 + n2-2) )
md <- mean(g2) - mean(g1)
semd <- sp * sqrt(1 / n1 + 1/n2)
rbind(
  md + c(-1, 1) * qt(.975, n1 + n2 - 2) * semd,
  t.test(g2, g1, paired = FALSE, var.equal = TRUE)$conf,
  t.test(g2, g1, paired = TRUE)$conf
)
```

```
##          [,1]  [,2]
## [1,] -0.2039 3.364
## [2,] -0.2039 3.364
## [3,]  0.7001 2.460
```

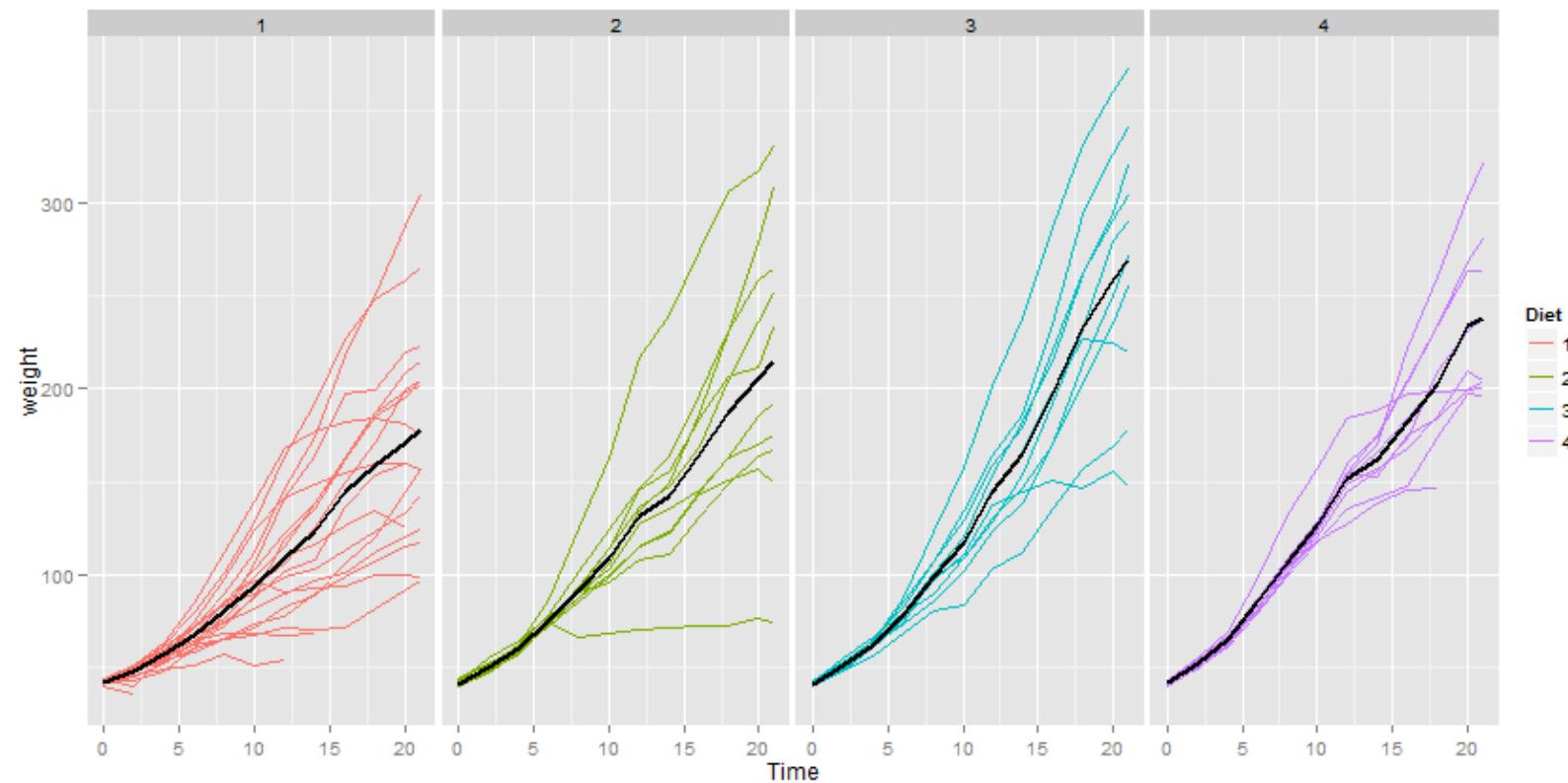
Grouped versus independent



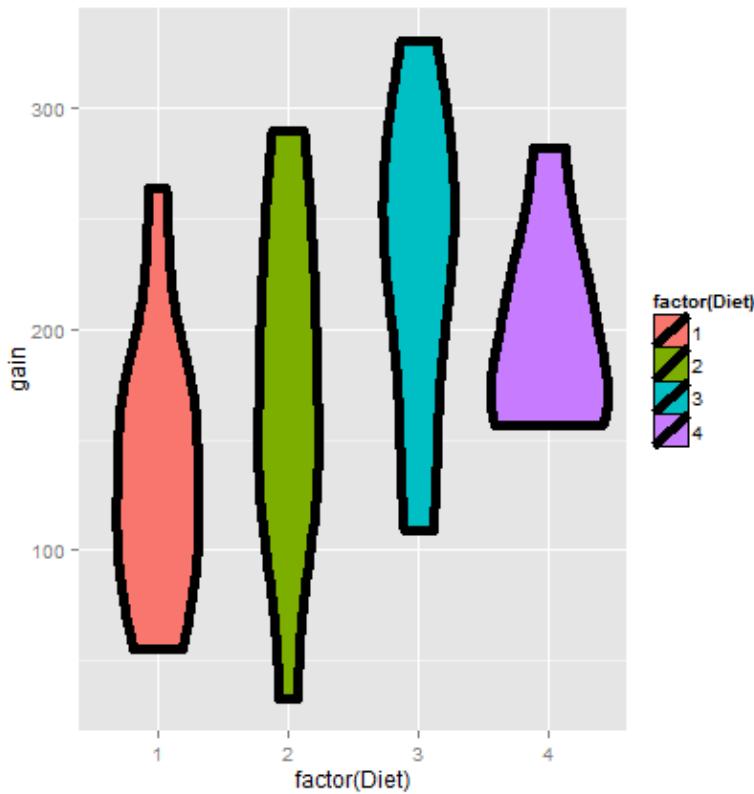
ChickWeight data in R

```
library(datasets); data(ChickWeight); library(reshape2)
##define weight gain or loss
wideCW <- dcast(ChickWeight, Diet + Chick ~ Time, value.var = "weight")
names(wideCW)[-c(1 : 2)] <- paste("time", names(wideCW)[-c(1 : 2)], sep = " ")
library(dplyr)
wideCW <- mutate(wideCW,
  gain = time21 - time0
)
```

Plotting the raw data



Weight gain by diet



Let's do a t interval

```
wideCW14 <- subset(wideCW, Diet %in% c(1, 4))
rbind(
t.test(gain ~ Diet, paired = FALSE, var.equal = TRUE, data = wideCW14)$conf,
t.test(gain ~ Diet, paired = FALSE, var.equal = FALSE, data = wideCW14)$conf
)
```

```
##          [,1]    [,2]
## [1,] -108.1 -14.81
## [2,] -104.7 -18.30
```

Unequal variances

- Under unequal variances

$$\bar{Y} - \bar{X} \pm t_{df} \times \left(\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y} \right)^{1/2}$$

where t_{df} is calculated with degrees of freedom

$$df = \frac{\left(S_x^2/n_x + S_y^2/n_y \right)^2}{\left(\frac{S_x^2}{n_x} \right)^2 / (n_x - 1) + \left(\frac{S_y^2}{n_y} \right)^2 / (n_y - 1)}$$

will be approximately a 95% interval

- This works really well
 - So when in doubt, just assume unequal variances

Example

- Comparing SBP for 8 oral contraceptive users versus 21 controls
- $\bar{X}_{OC} = 132.86 \text{ mmHg}$ with $s_{OC} = 15.34 \text{ mmHg}$
- $\bar{X}_C = 127.44 \text{ mmHg}$ with $s_C = 18.23 \text{ mmHg}$
- $df = 15.04, t_{15.04,.975} = 2.13$
- Interval

$$132.86 - 127.44 \pm 2.13 \left(\frac{15.34^2}{8} + \frac{18.23^2}{21} \right)^{1/2} = [-8.91, 19.75]$$

- In R, `t.test(..., var.equal = FALSE)`

Comparing other kinds of data

- For binomial data, there's lots of ways to compare two groups
 - Relative risk, risk difference, odds ratio.
 - Chi-squared tests, normal approximations, exact tests.
- For count data, there's also Chi-squared tests and exact tests.
- We'll leave the discussions for comparing groups of data for binary and count data until covering glms in the regression class.
- In addition, Mathematical Biostatistics Boot Camp 2 covers many special cases relevant to biostatistics.



Hypothesis testing

Statistical Inference

Brian Caffo, Jeff Leek, Roger Peng
Johns Hopkins Bloomberg School of Public Health

Hypothesis testing

- Hypothesis testing is concerned with making decisions using data
- A null hypothesis is specified that represents the status quo, usually labeled H_0
- The null hypothesis is assumed true and statistical evidence is required to reject it in favor of a research or alternative hypothesis

Example

- A respiratory disturbance index of more than 30 events / hour, say, is considered evidence of severe sleep disordered breathing (SDB).
- Suppose that in a sample of 100 overweight subjects with other risk factors for sleep disordered breathing at a sleep clinic, the mean RDI was 32 events / hour with a standard deviation of 10 events / hour.
- We might want to test the hypothesis that
 - $H_0 : \mu = 30$
 - $H_a : \mu > 30$
 - where μ is the population mean RDI.

Hypothesis testing

- The alternative hypotheses are typically of the form $<$, $>$ or \neq
- Note that there are four possible outcomes of our statistical decision process

TRUTH	DECIDE	RESULT
H_0	H_0	Correctly accept null
H_0	H_a	Type I error
H_a	H_a	Correctly reject null
H_a	H_0	Type II error

Discussion

- Consider a court of law; the null hypothesis is that the defendant is innocent
- We require a standard on the available evidence to reject the null hypothesis (convict)
- If we set a low standard, then we would increase the percentage of innocent people convicted (type I errors); however we would also increase the percentage of guilty people convicted (correctly rejecting the null)
- If we set a high standard, then we increase the the percentage of innocent people let free (correctly accepting the null) while we would also increase the percentage of guilty people let free (type II errors)

Example

- Consider our sleep example again
- A reasonable strategy would reject the null hypothesis if \bar{X} was larger than some constant, say C
- Typically, C is chosen so that the probability of a Type I error, α , is .05 (or some other relevant constant)
- α = Type I error rate = Probability of rejecting the null hypothesis when, in fact, the null hypothesis is correct

Example continued

- Standard error of the mean $10/\sqrt{100} = 1$
- Under H_0 $\bar{X} \sim N(30, 1)$
- We want to chose C so that the $P(\bar{X} > C; H_0)$ is 5%
- The 95th percentile of a normal distribution is 1.645 standard deviations from the mean
- If $C = 30 + 1 \times 1.645 = 31.645$
 - Then the probability that a $N(30, 1)$ is larger than it is 5%
 - So the rule "Reject H_0 when $\bar{X} \geq 31.645$ " has the property that the probability of rejection is 5% when H_0 is true (for the μ_0 , σ and n given)

Discussion

- In general we don't convert C back to the original scale
- We would just reject because the Z-score; which is how many standard errors the sample mean is above the hypothesized mean

$$\frac{32 - 30}{10/\sqrt{100}} = 2$$

is greater than 1.645

- Or, whenever $\sqrt{n}(\bar{X} - \mu_0)/s > Z_{1-\alpha}$

General rules

- The Z test for $H_0 : \mu = \mu_0$ versus
 - $H_1 : \mu < \mu_0$
 - $H_2 : \mu \neq \mu_0$
 - $H_3 : \mu > \mu_0$
- Test statistic $TS = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$
- Reject the null hypothesis when
 - $TS \leq Z_\alpha = -Z_{1-\alpha}$
 - $|TS| \geq Z_{1-\alpha/2}$
 - $TS \geq Z_{1-\alpha}$

Notes

- We have fixed α to be low, so if we reject H_0 (either our model is wrong) or there is a low probability that we have made an error
- We have not fixed the probability of a type II error, β ; therefore we tend to say ``Fail to reject H_0 '' rather than accepting H_0
- Statistical significance is no the same as scientific significance
- The region of TS values for which you reject H_0 is called the rejection region

More notes

- The Z test requires the assumptions of the CLT and for n to be large enough for it to apply
- If n is small, then a Gossett's T test is performed exactly in the same way, with the normal quantiles replaced by the appropriate Student's T quantiles and $n - 1$ df
- The probability of rejecting the null hypothesis when it is false is called *power*
- Power is used a lot to calculate sample sizes for experiments

Example reconsidered

- Consider our example again. Suppose that $n = 16$ (rather than 100)
- The statistic

$$\frac{\bar{X} - 30}{s/\sqrt{16}}$$

follows a T distribution with 15 df under H_0

- Under H_0 , the probability that it is larger than the 95th percentile of the T distribution is 5%
- The 95th percentile of the T distribution with 15 df is 1.7531 (obtained via `qt(.95, 15)`)
- So that our test statistic is now $\sqrt{16}(32 - 30)/10 = 0.8$
- We now fail to reject.

Two sided tests

- Suppose that we would reject the null hypothesis if in fact the mean was too large or too small
- That is, we want to test the alternative $H_a : \mu \neq 30$
- We will reject if the test statistic, 0.8, is either too large or too small
- Then we want the probability of rejecting under the null to be 5%, split equally as 2.5% in the upper tail and 2.5% in the lower tail
- Thus we reject if our test statistic is larger than $qt(.975, 15)$ or smaller than $qt(.025, 15)$
 - This is the same as saying: reject if the absolute value of our statistic is larger than $qt(0.975, 15) = 2.1314$
 - So we fail to reject the two sided test as well
 - (If you fail to reject the one sided test, you know that you will fail to reject the two sided)

T test in R

```
library(UsingR); data(father.son)
t.test(father.son$sheight - father.son$fheight)
```

```
>
> One Sample t-test
>
> data: father.son$sheight - father.son$fheight
> t = 11.79, df = 1077, p-value < 2.2e-16
> alternative hypothesis: true mean is not equal to 0
> 95 percent confidence interval:
>  0.831 1.163
> sample estimates:
> mean of x
>      0.997
```

Connections with confidence intervals

- Consider testing $H_0 : \mu = \mu_0$ versus $H_a : \mu \neq \mu_0$
- Take the set of all possible values for which you fail to reject H_0 , this set is a $(1 - \alpha)100\%$ confidence interval for μ
- The same works in reverse; if a $(1 - \alpha)100\%$ interval contains μ_0 , then we *fail to* reject H_0

Two group intervals

- First, now you know how to do two group T tests since we already covered indepedent group T intervals
- Rejection rules are the same
- Test $H_0 : \mu_1 = \mu_2$
- Let's just go through an example

chickWeight data

Recall that we reformatted this data

```
library(datasets); data(ChickWeight); library(reshape2)
##define weight gain or loss
wideCW <- dcast(ChickWeight, Diet + Chick ~ Time, value.var = "weight")
names(wideCW)[-(1 : 2)] <- paste("time", names(wideCW)[-(1 : 2)], sep = " ")
library(dplyr)
wideCW <- mutate(wideCW,
  gain = time21 - time0
)
```

Unequal variance T test comparing diets 1 and 4

```
wideCW14 <- subset(wideCW, Diet %in% c(1, 4))
t.test(gain ~ Diet, paired = FALSE,
       var.equal = TRUE, data = wideCW14)
```

```
>
> Two Sample t-test
>
> data: gain by Diet
> t = -2.725, df = 23, p-value = 0.01207
> alternative hypothesis: true difference in means is not equal to 0
> 95 percent confidence interval:
> -108.15 -14.81
> sample estimates:
> mean in group 1 mean in group 4
>           136.2          197.7
```

Exact binomial test

- Recall this problem, *Suppose a friend has 8 children, 7 of which are girls and none are twins*
- Perform the relevant hypothesis test. $H_0 : p = 0.5$ $H_a : p > 0.5$
 - What is the relevant rejection region so that the probability of rejecting is (less than) 5%?

REJECTION REGION	TYPE I ERROR RATE
[0 : 8]	1
[1 : 8]	0.9961
[2 : 8]	0.9648
[3 : 8]	0.8555
[4 : 8]	0.6367
[5 : 8]	0.3633
[6 : 8]	0.1445
[7 : 8]	0.0352
[8 : 8]	0.0039

Notes

- It's impossible to get an exact 5% level test for this case due to the discreteness of the binomial.
 - The closest is the rejection region [7 : 8]
 - Any alpha level lower than 0.0039 is not attainable.
- For larger sample sizes, we could do a normal approximation, but you already knew this.
- Two sided test isn't obvious.
 - Given a way to do two sided tests, we could take the set of values of p_0 for which we fail to reject to get an exact binomial confidence interval (called the Clopper/Pearson interval, BTW)
- For these problems, people always create a P-value (next lecture) rather than computing the rejection region.



Hypothesis testing

Statistical Inference

Brian Caffo, Jeff Leek, Roger Peng
Johns Hopkins Bloomberg School of Public Health

Hypothesis testing

- Hypothesis testing is concerned with making decisions using data
- A null hypothesis is specified that represents the status quo, usually labeled H_0
- The null hypothesis is assumed true and statistical evidence is required to reject it in favor of a research or alternative hypothesis

Example

- A respiratory disturbance index of more than 30 events / hour, say, is considered evidence of severe sleep disordered breathing (SDB).
- Suppose that in a sample of 100 overweight subjects with other risk factors for sleep disordered breathing at a sleep clinic, the mean RDI was 32 events / hour with a standard deviation of 10 events / hour.
- We might want to test the hypothesis that
 - $H_0 : \mu = 30$
 - $H_a : \mu > 30$
 - where μ is the population mean RDI.

Hypothesis testing

- The alternative hypotheses are typically of the form $<$, $>$ or \neq
- Note that there are four possible outcomes of our statistical decision process

TRUTH	DECIDE	RESULT
H_0	H_0	Correctly accept null
H_0	H_a	Type I error
H_a	H_a	Correctly reject null
H_a	H_0	Type II error

Discussion

- Consider a court of law; the null hypothesis is that the defendant is innocent
- We require a standard on the available evidence to reject the null hypothesis (convict)
- If we set a low standard, then we would increase the percentage of innocent people convicted (type I errors); however we would also increase the percentage of guilty people convicted (correctly rejecting the null)
- If we set a high standard, then we increase the the percentage of innocent people let free (correctly accepting the null) while we would also increase the percentage of guilty people let free (type II errors)

Example

- Consider our sleep example again
- A reasonable strategy would reject the null hypothesis if \bar{X} was larger than some constant, say C
- Typically, C is chosen so that the probability of a Type I error, α , is .05 (or some other relevant constant)
- α = Type I error rate = Probability of rejecting the null hypothesis when, in fact, the null hypothesis is correct

Example continued

- Standard error of the mean $10/\sqrt{100} = 1$
- Under H_0 $\bar{X} \sim N(30, 1)$
- We want to chose C so that the $P(\bar{X} > C; H_0)$ is 5%
- The 95th percentile of a normal distribution is 1.645 standard deviations from the mean
- If $C = 30 + 1 \times 1.645 = 31.645$
 - Then the probability that a $N(30, 1)$ is larger than it is 5%
 - So the rule "Reject H_0 when $\bar{X} \geq 31.645$ " has the property that the probability of rejection is 5% when H_0 is true (for the μ_0 , σ and n given)

Discussion

- In general we don't convert C back to the original scale
- We would just reject because the Z-score; which is how many standard errors the sample mean is above the hypothesized mean

$$\frac{32 - 30}{10/\sqrt{100}} = 2$$

is greater than 1.645

- Or, whenever $\sqrt{n}(\bar{X} - \mu_0)/s > Z_{1-\alpha}$

General rules

- The Z test for $H_0 : \mu = \mu_0$ versus
 - $H_1 : \mu < \mu_0$
 - $H_2 : \mu \neq \mu_0$
 - $H_3 : \mu > \mu_0$
- Test statistic $TS = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$
- Reject the null hypothesis when
 - $TS \leq Z_\alpha = -Z_{1-\alpha}$
 - $|TS| \geq Z_{1-\alpha/2}$
 - $TS \geq Z_{1-\alpha}$

Notes

- We have fixed α to be low, so if we reject H_0 (either our model is wrong) or there is a low probability that we have made an error
- We have not fixed the probability of a type II error, β ; therefore we tend to say ``Fail to reject H_0 '' rather than accepting H_0
- Statistical significance is no the same as scientific significance
- The region of TS values for which you reject H_0 is called the rejection region

More notes

- The Z test requires the assumptions of the CLT and for n to be large enough for it to apply
- If n is small, then a Gossett's T test is performed exactly in the same way, with the normal quantiles replaced by the appropriate Student's T quantiles and $n - 1$ df
- The probability of rejecting the null hypothesis when it is false is called *power*
- Power is used a lot to calculate sample sizes for experiments

Example reconsidered

- Consider our example again. Suppose that $n = 16$ (rather than 100)
- The statistic

$$\frac{\bar{X} - 30}{s/\sqrt{16}}$$

follows a T distribution with 15 df under H_0

- Under H_0 , the probability that it is larger than the 95th percentile of the T distribution is 5%
- The 95th percentile of the T distribution with 15 df is 1.7531 (obtained via `qt(.95, 15)`)
- So that our test statistic is now $\sqrt{16}(32 - 30)/10 = 0.8$
- We now fail to reject.

Two sided tests

- Suppose that we would reject the null hypothesis if in fact the mean was too large or too small
- That is, we want to test the alternative $H_a : \mu \neq 30$
- We will reject if the test statistic, 0.8, is either too large or too small
- Then we want the probability of rejecting under the null to be 5%, split equally as 2.5% in the upper tail and 2.5% in the lower tail
- Thus we reject if our test statistic is larger than $qt(.975, 15)$ or smaller than $qt(.025, 15)$
 - This is the same as saying: reject if the absolute value of our statistic is larger than $qt(0.975, 15) = 2.1314$
 - So we fail to reject the two sided test as well
 - (If you fail to reject the one sided test, you know that you will fail to reject the two sided)

T test in R

```
library(UsingR); data(father.son)
t.test(father.son$sheight - father.son$fheight)
```

```
>
> One Sample t-test
>
> data: father.son$sheight - father.son$fheight
> t = 11.79, df = 1077, p-value < 2.2e-16
> alternative hypothesis: true mean is not equal to 0
> 95 percent confidence interval:
>  0.831 1.163
> sample estimates:
> mean of x
>      0.997
```

Connections with confidence intervals

- Consider testing $H_0 : \mu = \mu_0$ versus $H_a : \mu \neq \mu_0$
- Take the set of all possible values for which you fail to reject H_0 , this set is a $(1 - \alpha)100\%$ confidence interval for μ
- The same works in reverse; if a $(1 - \alpha)100\%$ interval contains μ_0 , then we *fail to* reject H_0

Two group intervals

- First, now you know how to do two group T tests since we already covered indepedent group T intervals
- Rejection rules are the same
- Test $H_0 : \mu_1 = \mu_2$
- Let's just go through an example

chickWeight data

Recall that we reformatted this data

```
library(datasets); data(ChickWeight); library(reshape2)
##define weight gain or loss
wideCW <- dcast(ChickWeight, Diet + Chick ~ Time, value.var = "weight")
names(wideCW)[-(1 : 2)] <- paste("time", names(wideCW)[-(1 : 2)], sep = " ")
library(dplyr)
wideCW <- mutate(wideCW,
  gain = time21 - time0
)
```

Unequal variance T test comparing diets 1 and 4

```
wideCW14 <- subset(wideCW, Diet %in% c(1, 4))
t.test(gain ~ Diet, paired = FALSE,
       var.equal = TRUE, data = wideCW14)
```

```
>
> Two Sample t-test
>
> data: gain by Diet
> t = -2.725, df = 23, p-value = 0.01207
> alternative hypothesis: true difference in means is not equal to 0
> 95 percent confidence interval:
> -108.15 -14.81
> sample estimates:
> mean in group 1 mean in group 4
>           136.2          197.7
```

Exact binomial test

- Recall this problem, *Suppose a friend has 8 children, 7 of which are girls and none are twins*
- Perform the relevant hypothesis test. $H_0 : p = 0.5$ $H_a : p > 0.5$
 - What is the relevant rejection region so that the probability of rejecting is (less than) 5%?

REJECTION REGION	TYPE I ERROR RATE
[0 : 8]	1
[1 : 8]	0.9961
[2 : 8]	0.9648
[3 : 8]	0.8555
[4 : 8]	0.6367
[5 : 8]	0.3633
[6 : 8]	0.1445
[7 : 8]	0.0352
[8 : 8]	0.0039

Notes

- It's impossible to get an exact 5% level test for this case due to the discreteness of the binomial.
 - The closest is the rejection region [7 : 8]
 - Any alpha level lower than 0.0039 is not attainable.
- For larger sample sizes, we could do a normal approximation, but you already knew this.
- Two sided test isn't obvious.
 - Given a way to do two sided tests, we could take the set of values of p_0 for which we fail to reject to get an exact binomial confidence interval (called the Clopper/Pearson interval, BTW)
- For these problems, people always create a P-value (next lecture) rather than computing the rejection region.



Hypothesis testing

Statistical Inference

Brian Caffo, Jeff Leek, Roger Peng
Johns Hopkins Bloomberg School of Public Health

Hypothesis testing

- Hypothesis testing is concerned with making decisions using data
- A null hypothesis is specified that represents the status quo, usually labeled H_0
- The null hypothesis is assumed true and statistical evidence is required to reject it in favor of a research or alternative hypothesis

Example

- A respiratory disturbance index of more than 30 events / hour, say, is considered evidence of severe sleep disordered breathing (SDB).
- Suppose that in a sample of 100 overweight subjects with other risk factors for sleep disordered breathing at a sleep clinic, the mean RDI was 32 events / hour with a standard deviation of 10 events / hour.
- We might want to test the hypothesis that
 - $H_0 : \mu = 30$
 - $H_a : \mu > 30$
 - where μ is the population mean RDI.

Hypothesis testing

- The alternative hypotheses are typically of the form $<$, $>$ or \neq
- Note that there are four possible outcomes of our statistical decision process

TRUTH	DECIDE	RESULT
H_0	H_0	Correctly accept null
H_0	H_a	Type I error
H_a	H_a	Correctly reject null
H_a	H_0	Type II error

Discussion

- Consider a court of law; the null hypothesis is that the defendant is innocent
- We require a standard on the available evidence to reject the null hypothesis (convict)
- If we set a low standard, then we would increase the percentage of innocent people convicted (type I errors); however we would also increase the percentage of guilty people convicted (correctly rejecting the null)
- If we set a high standard, then we increase the the percentage of innocent people let free (correctly accepting the null) while we would also increase the percentage of guilty people let free (type II errors)

Example

- Consider our sleep example again
- A reasonable strategy would reject the null hypothesis if \bar{X} was larger than some constant, say C
- Typically, C is chosen so that the probability of a Type I error, α , is .05 (or some other relevant constant)
- α = Type I error rate = Probability of rejecting the null hypothesis when, in fact, the null hypothesis is correct

Example continued

- Standard error of the mean $10/\sqrt{100} = 1$
- Under H_0 $\bar{X} \sim N(30, 1)$
- We want to chose C so that the $P(\bar{X} > C; H_0)$ is 5%
- The 95th percentile of a normal distribution is 1.645 standard deviations from the mean
- If $C = 30 + 1 \times 1.645 = 31.645$
 - Then the probability that a $N(30, 1)$ is larger than it is 5%
 - So the rule "Reject H_0 when $\bar{X} \geq 31.645$ " has the property that the probability of rejection is 5% when H_0 is true (for the μ_0 , σ and n given)

Discussion

- In general we don't convert C back to the original scale
- We would just reject because the Z-score; which is how many standard errors the sample mean is above the hypothesized mean

$$\frac{32 - 30}{10/\sqrt{100}} = 2$$

is greater than 1.645

- Or, whenever $\sqrt{n}(\bar{X} - \mu_0)/s > Z_{1-\alpha}$

General rules

- The Z test for $H_0 : \mu = \mu_0$ versus
 - $H_1 : \mu < \mu_0$
 - $H_2 : \mu \neq \mu_0$
 - $H_3 : \mu > \mu_0$
- Test statistic $TS = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$
- Reject the null hypothesis when
 - $TS \leq Z_\alpha = -Z_{1-\alpha}$
 - $|TS| \geq Z_{1-\alpha/2}$
 - $TS \geq Z_{1-\alpha}$

Notes

- We have fixed α to be low, so if we reject H_0 (either our model is wrong) or there is a low probability that we have made an error
- We have not fixed the probability of a type II error, β ; therefore we tend to say ``Fail to reject H_0 '' rather than accepting H_0
- Statistical significance is no the same as scientific significance
- The region of TS values for which you reject H_0 is called the rejection region

More notes

- The Z test requires the assumptions of the CLT and for n to be large enough for it to apply
- If n is small, then a Gossett's T test is performed exactly in the same way, with the normal quantiles replaced by the appropriate Student's T quantiles and $n - 1$ df
- The probability of rejecting the null hypothesis when it is false is called *power*
- Power is used a lot to calculate sample sizes for experiments

Example reconsidered

- Consider our example again. Suppose that $n = 16$ (rather than 100)
- The statistic

$$\frac{\bar{X} - 30}{s/\sqrt{16}}$$

follows a T distribution with 15 df under H_0

- Under H_0 , the probability that it is larger than the 95th percentile of the T distribution is 5%
- The 95th percentile of the T distribution with 15 df is 1.7531 (obtained via `qt(.95, 15)`)
- So that our test statistic is now $\sqrt{16}(32 - 30)/10 = 0.8$
- We now fail to reject.

Two sided tests

- Suppose that we would reject the null hypothesis if in fact the mean was too large or too small
- That is, we want to test the alternative $H_a : \mu \neq 30$
- We will reject if the test statistic, 0.8, is either too large or too small
- Then we want the probability of rejecting under the null to be 5%, split equally as 2.5% in the upper tail and 2.5% in the lower tail
- Thus we reject if our test statistic is larger than $qt(.975, 15)$ or smaller than $qt(.025, 15)$
 - This is the same as saying: reject if the absolute value of our statistic is larger than $qt(0.975, 15) = 2.1314$
 - So we fail to reject the two sided test as well
 - (If you fail to reject the one sided test, you know that you will fail to reject the two sided)

T test in R

```
library(UsingR); data(father.son)
t.test(father.son$sheight - father.son$fheight)
```

```
>
> One Sample t-test
>
> data: father.son$sheight - father.son$fheight
> t = 11.79, df = 1077, p-value < 2.2e-16
> alternative hypothesis: true mean is not equal to 0
> 95 percent confidence interval:
>  0.831 1.163
> sample estimates:
> mean of x
>      0.997
```

Connections with confidence intervals

- Consider testing $H_0 : \mu = \mu_0$ versus $H_a : \mu \neq \mu_0$
- Take the set of all possible values for which you fail to reject H_0 , this set is a $(1 - \alpha)100\%$ confidence interval for μ
- The same works in reverse; if a $(1 - \alpha)100\%$ interval contains μ_0 , then we *fail to* reject H_0

Two group intervals

- First, now you know how to do two group T tests since we already covered indepedent group T intervals
- Rejection rules are the same
- Test $H_0 : \mu_1 = \mu_2$
- Let's just go through an example

chickWeight data

Recall that we reformatted this data

```
library(datasets); data(ChickWeight); library(reshape2)
##define weight gain or loss
wideCW <- dcast(ChickWeight, Diet + Chick ~ Time, value.var = "weight")
names(wideCW)[-(1 : 2)] <- paste("time", names(wideCW)[-(1 : 2)], sep = " ")
library(dplyr)
wideCW <- mutate(wideCW,
  gain = time21 - time0
)
```

Unequal variance T test comparing diets 1 and 4

```
wideCW14 <- subset(wideCW, Diet %in% c(1, 4))
t.test(gain ~ Diet, paired = FALSE,
       var.equal = TRUE, data = wideCW14)
```

```
>
> Two Sample t-test
>
> data: gain by Diet
> t = -2.725, df = 23, p-value = 0.01207
> alternative hypothesis: true difference in means is not equal to 0
> 95 percent confidence interval:
> -108.15 -14.81
> sample estimates:
> mean in group 1 mean in group 4
>           136.2          197.7
```

Exact binomial test

- Recall this problem, *Suppose a friend has 8 children, 7 of which are girls and none are twins*
- Perform the relevant hypothesis test. $H_0 : p = 0.5$ $H_a : p > 0.5$
 - What is the relevant rejection region so that the probability of rejecting is (less than) 5%?

REJECTION REGION	TYPE I ERROR RATE
[0 : 8]	1
[1 : 8]	0.9961
[2 : 8]	0.9648
[3 : 8]	0.8555
[4 : 8]	0.6367
[5 : 8]	0.3633
[6 : 8]	0.1445
[7 : 8]	0.0352
[8 : 8]	0.0039

Notes

- It's impossible to get an exact 5% level test for this case due to the discreteness of the binomial.
 - The closest is the rejection region [7 : 8]
 - Any alpha level lower than 0.0039 is not attainable.
- For larger sample sizes, we could do a normal approximation, but you already knew this.
- Two sided test isn't obvious.
 - Given a way to do two sided tests, we could take the set of values of p_0 for which we fail to reject to get an exact binomial confidence interval (called the Clopper/Pearson interval, BTW)
- For these problems, people always create a P-value (next lecture) rather than computing the rejection region.



Hypothesis testing

Statistical Inference

Brian Caffo, Jeff Leek, Roger Peng
Johns Hopkins Bloomberg School of Public Health

Hypothesis testing

- Hypothesis testing is concerned with making decisions using data
- A null hypothesis is specified that represents the status quo, usually labeled H_0
- The null hypothesis is assumed true and statistical evidence is required to reject it in favor of a research or alternative hypothesis

Example

- A respiratory disturbance index of more than 30 events / hour, say, is considered evidence of severe sleep disordered breathing (SDB).
- Suppose that in a sample of 100 overweight subjects with other risk factors for sleep disordered breathing at a sleep clinic, the mean RDI was 32 events / hour with a standard deviation of 10 events / hour.
- We might want to test the hypothesis that
 - $H_0 : \mu = 30$
 - $H_a : \mu > 30$
 - where μ is the population mean RDI.

Hypothesis testing

- The alternative hypotheses are typically of the form $<$, $>$ or \neq
- Note that there are four possible outcomes of our statistical decision process

TRUTH	DECIDE	RESULT
H_0	H_0	Correctly accept null
H_0	H_a	Type I error
H_a	H_a	Correctly reject null
H_a	H_0	Type II error

Discussion

- Consider a court of law; the null hypothesis is that the defendant is innocent
- We require a standard on the available evidence to reject the null hypothesis (convict)
- If we set a low standard, then we would increase the percentage of innocent people convicted (type I errors); however we would also increase the percentage of guilty people convicted (correctly rejecting the null)
- If we set a high standard, then we increase the the percentage of innocent people let free (correctly accepting the null) while we would also increase the percentage of guilty people let free (type II errors)

Example

- Consider our sleep example again
- A reasonable strategy would reject the null hypothesis if \bar{X} was larger than some constant, say C
- Typically, C is chosen so that the probability of a Type I error, α , is .05 (or some other relevant constant)
- α = Type I error rate = Probability of rejecting the null hypothesis when, in fact, the null hypothesis is correct

Example continued

- Standard error of the mean $10/\sqrt{100} = 1$
- Under H_0 $\bar{X} \sim N(30, 1)$
- We want to chose C so that the $P(\bar{X} > C; H_0)$ is 5%
- The 95th percentile of a normal distribution is 1.645 standard deviations from the mean
- If $C = 30 + 1 \times 1.645 = 31.645$
 - Then the probability that a $N(30, 1)$ is larger than it is 5%
 - So the rule "Reject H_0 when $\bar{X} \geq 31.645$ " has the property that the probability of rejection is 5% when H_0 is true (for the μ_0 , σ and n given)

Discussion

- In general we don't convert C back to the original scale
- We would just reject because the Z-score; which is how many standard errors the sample mean is above the hypothesized mean

$$\frac{32 - 30}{10/\sqrt{100}} = 2$$

is greater than 1.645

- Or, whenever $\sqrt{n}(\bar{X} - \mu_0)/s > Z_{1-\alpha}$

General rules

- The Z test for $H_0 : \mu = \mu_0$ versus
 - $H_1 : \mu < \mu_0$
 - $H_2 : \mu \neq \mu_0$
 - $H_3 : \mu > \mu_0$
- Test statistic $TS = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$
- Reject the null hypothesis when
 - $TS \leq Z_\alpha = -Z_{1-\alpha}$
 - $|TS| \geq Z_{1-\alpha/2}$
 - $TS \geq Z_{1-\alpha}$

Notes

- We have fixed α to be low, so if we reject H_0 (either our model is wrong) or there is a low probability that we have made an error
- We have not fixed the probability of a type II error, β ; therefore we tend to say ``Fail to reject H_0 '' rather than accepting H_0
- Statistical significance is no the same as scientific significance
- The region of TS values for which you reject H_0 is called the rejection region

More notes

- The Z test requires the assumptions of the CLT and for n to be large enough for it to apply
- If n is small, then a Gossett's T test is performed exactly in the same way, with the normal quantiles replaced by the appropriate Student's T quantiles and $n - 1$ df
- The probability of rejecting the null hypothesis when it is false is called *power*
- Power is used a lot to calculate sample sizes for experiments

Example reconsidered

- Consider our example again. Suppose that $n = 16$ (rather than 100)
- The statistic

$$\frac{\bar{X} - 30}{s/\sqrt{16}}$$

follows a T distribution with 15 df under H_0

- Under H_0 , the probability that it is larger than the 95th percentile of the T distribution is 5%
- The 95th percentile of the T distribution with 15 df is 1.7531 (obtained via `qt(.95, 15)`)
- So that our test statistic is now $\sqrt{16}(32 - 30)/10 = 0.8$
- We now fail to reject.

Two sided tests

- Suppose that we would reject the null hypothesis if in fact the mean was too large or too small
- That is, we want to test the alternative $H_a : \mu \neq 30$
- We will reject if the test statistic, 0.8, is either too large or too small
- Then we want the probability of rejecting under the null to be 5%, split equally as 2.5% in the upper tail and 2.5% in the lower tail
- Thus we reject if our test statistic is larger than $qt(.975, 15)$ or smaller than $qt(.025, 15)$
 - This is the same as saying: reject if the absolute value of our statistic is larger than $qt(0.975, 15) = 2.1314$
 - So we fail to reject the two sided test as well
 - (If you fail to reject the one sided test, you know that you will fail to reject the two sided)

T test in R

```
library(UsingR); data(father.son)
t.test(father.son$sheight - father.son$fheight)
```

```
>
> One Sample t-test
>
> data: father.son$sheight - father.son$fheight
> t = 11.79, df = 1077, p-value < 2.2e-16
> alternative hypothesis: true mean is not equal to 0
> 95 percent confidence interval:
>  0.831 1.163
> sample estimates:
> mean of x
>      0.997
```

Connections with confidence intervals

- Consider testing $H_0 : \mu = \mu_0$ versus $H_a : \mu \neq \mu_0$
- Take the set of all possible values for which you fail to reject H_0 , this set is a $(1 - \alpha)100\%$ confidence interval for μ
- The same works in reverse; if a $(1 - \alpha)100\%$ interval contains μ_0 , then we *fail to* reject H_0

Two group intervals

- First, now you know how to do two group T tests since we already covered indepedent group T intervals
- Rejection rules are the same
- Test $H_0 : \mu_1 = \mu_2$
- Let's just go through an example

chickWeight data

Recall that we reformatted this data

```
library(datasets); data(ChickWeight); library(reshape2)
##define weight gain or loss
wideCW <- dcast(ChickWeight, Diet + Chick ~ Time, value.var = "weight")
names(wideCW)[-(1 : 2)] <- paste("time", names(wideCW)[-(1 : 2)], sep = " ")
library(dplyr)
wideCW <- mutate(wideCW,
  gain = time21 - time0
)
```

Unequal variance T test comparing diets 1 and 4

```
wideCW14 <- subset(wideCW, Diet %in% c(1, 4))
t.test(gain ~ Diet, paired = FALSE,
       var.equal = TRUE, data = wideCW14)
```

```
>
> Two Sample t-test
>
> data: gain by Diet
> t = -2.725, df = 23, p-value = 0.01207
> alternative hypothesis: true difference in means is not equal to 0
> 95 percent confidence interval:
> -108.15 -14.81
> sample estimates:
> mean in group 1 mean in group 4
>           136.2          197.7
```

Exact binomial test

- Recall this problem, *Suppose a friend has 8 children, 7 of which are girls and none are twins*
- Perform the relevant hypothesis test. $H_0 : p = 0.5$ $H_a : p > 0.5$
 - What is the relevant rejection region so that the probability of rejecting is (less than) 5%?

REJECTION REGION	TYPE I ERROR RATE
[0 : 8]	1
[1 : 8]	0.9961
[2 : 8]	0.9648
[3 : 8]	0.8555
[4 : 8]	0.6367
[5 : 8]	0.3633
[6 : 8]	0.1445
[7 : 8]	0.0352
[8 : 8]	0.0039

Notes

- It's impossible to get an exact 5% level test for this case due to the discreteness of the binomial.
 - The closest is the rejection region [7 : 8]
 - Any alpha level lower than 0.0039 is not attainable.
- For larger sample sizes, we could do a normal approximation, but you already knew this.
- Two sided test isn't obvious.
 - Given a way to do two sided tests, we could take the set of values of p_0 for which we fail to reject to get an exact binomial confidence interval (called the Clopper/Pearson interval, BTW)
- For these problems, people always create a P-value (next lecture) rather than computing the rejection region.



P-values

Statistical inference

Brian Caffo, Jeffrey Leek, Roger Peng
Johns Hopkins Bloomberg School of Public Health

P-values

- Most common measure of "statistical significance"
- Their ubiquity, along with concern over their interpretation and use makes them controversial among statisticians
 - <http://warnercnr.colostate.edu/~anderson/thompson1.html>
 - Also see *Statistical Evidence: A Likelihood Paradigm* by Richard Royall
 - *Toward Evidence-Based Medical Statistics. 1: The P Value Fallacy* by Steve Goodman
 - The hilariously titled: *The Earth is Round ($p < .05$)* by Cohen.
- Some positive comments
 - [simply statistics](#)
 - [normal deviate](#)
 - [Error statistics](#)

What is a P-value?

Idea: Suppose nothing is going on - how unusual is it to see the estimate we got?

Approach:

1. Define the hypothetical distribution of a data summary (statistic) when "nothing is going on" (*null hypothesis*)
2. Calculate the summary/statistic with the data we have (*test statistic*)
3. Compare what we calculated to our hypothetical distribution and see if the value is "extreme" (*p-value*)

P-values

- The P-value is the probability under the null hypothesis of obtaining evidence as extreme or more extreme than would be observed by chance alone
- If the P-value is small, then either H_0 is true and we have observed a rare event or H_0 is false
- In our example the T statistic was 0.8.
 - What's the probability of getting a T statistic as large as 0.8?

```
pt(0.8, 15, lower.tail = FALSE)
```

```
## [1] 0.2181
```

- Therefore, the probability of seeing evidence as extreme or more extreme than that actually obtained under H_0 is 0.2181

The attained significance level

- Our test statistic was 2 for $H_0 : \mu_0 = 30$ versus $H_a : \mu > 30$.
- Notice that we rejected the one sided test when $\alpha = 0.05$, would we reject if $\alpha = 0.01$, how about 0.001 ?
- The smallest value for alpha that you still reject the null hypothesis is called the *attained significance level*
- This is equivalent, but philosophically a little different from, the *P-value*

Notes

- By reporting a P-value the reader can perform the hypothesis test at whatever α level he or she chooses
- If the P-value is less than α you reject the null hypothesis
- For two sided hypothesis test, double the smaller of the two one sided hypothesis test Pvalues

Revisiting an earlier example

- Suppose a friend has 8 children, 7 of which are girls and none are twins
- If each gender has an independent 50% probability for each birth, what's the probability of getting 7 or more girls out of 8 births?

```
choose(8, 7) * 0.5^8 + choose(8, 8) * 0.5^8
```

```
## [1] 0.03516
```

```
pbinom(6, size = 8, prob = 0.5, lower.tail = FALSE)
```

```
## [1] 0.03516
```

Poisson example

- Suppose that a hospital has an infection rate of 10 infections per 100 person/days at risk (rate of 0.1) during the last monitoring period.
- Assume that an infection rate of 0.05 is an important benchmark.
- Given the model, could the observed rate being larger than 0.05 be attributed to chance?
- Under $H_0 : \lambda = 0.05$ so that $\lambda_0 100 = 5$
- Consider $H_a : \lambda > 0.05$.

```
ppois(9, 5, lower.tail = FALSE)
```

```
## [1] 0.03183
```



P-values

Statistical inference

Brian Caffo, Jeffrey Leek, Roger Peng
Johns Hopkins Bloomberg School of Public Health

P-values

- Most common measure of "statistical significance"
- Their ubiquity, along with concern over their interpretation and use makes them controversial among statisticians
 - <http://warnercnr.colostate.edu/~anderson/thompson1.html>
 - Also see *Statistical Evidence: A Likelihood Paradigm* by Richard Royall
 - *Toward Evidence-Based Medical Statistics. 1: The P Value Fallacy* by Steve Goodman
 - The hilariously titled: *The Earth is Round ($p < .05$)* by Cohen.
- Some positive comments
 - [simply statistics](#)
 - [normal deviate](#)
 - [Error statistics](#)

What is a P-value?

Idea: Suppose nothing is going on - how unusual is it to see the estimate we got?

Approach:

1. Define the hypothetical distribution of a data summary (statistic) when "nothing is going on" (*null hypothesis*)
2. Calculate the summary/statistic with the data we have (*test statistic*)
3. Compare what we calculated to our hypothetical distribution and see if the value is "extreme" (*p-value*)

P-values

- The P-value is the probability under the null hypothesis of obtaining evidence as extreme or more extreme than would be observed by chance alone
- If the P-value is small, then either H_0 is true and we have observed a rare event or H_0 is false
- In our example the T statistic was 0.8.
 - What's the probability of getting a T statistic as large as 0.8?

```
pt(0.8, 15, lower.tail = FALSE)
```

```
## [1] 0.2181
```

- Therefore, the probability of seeing evidence as extreme or more extreme than that actually obtained under H_0 is 0.2181

The attained significance level

- Our test statistic was 2 for $H_0 : \mu_0 = 30$ versus $H_a : \mu > 30$.
- Notice that we rejected the one sided test when $\alpha = 0.05$, would we reject if $\alpha = 0.01$, how about 0.001 ?
- The smallest value for alpha that you still reject the null hypothesis is called the *attained significance level*
- This is equivalent, but philosophically a little different from, the *P-value*

Notes

- By reporting a P-value the reader can perform the hypothesis test at whatever α level he or she chooses
- If the P-value is less than α you reject the null hypothesis
- For two sided hypothesis test, double the smaller of the two one sided hypothesis test Pvalues

Revisiting an earlier example

- Suppose a friend has 8 children, 7 of which are girls and none are twins
- If each gender has an independent 50% probability for each birth, what's the probability of getting 7 or more girls out of 8 births?

```
choose(8, 7) * 0.5^8 + choose(8, 8) * 0.5^8
```

```
## [1] 0.03516
```

```
pbinom(6, size = 8, prob = 0.5, lower.tail = FALSE)
```

```
## [1] 0.03516
```

Poisson example

- Suppose that a hospital has an infection rate of 10 infections per 100 person/days at risk (rate of 0.1) during the last monitoring period.
- Assume that an infection rate of 0.05 is an important benchmark.
- Given the model, could the observed rate being larger than 0.05 be attributed to chance?
- Under $H_0 : \lambda = 0.05$ so that $\lambda_0 100 = 5$
- Consider $H_a : \lambda > 0.05$.

```
ppois(9, 5, lower.tail = FALSE)
```

```
## [1] 0.03183
```