STA258H5
University of Toronto Mississauga

Al Nosedal, Asal Aslemand and Omid Jazi

Winter 2023

Hypothesis Tests

- For the Difference Between Two Means
- For the Difference Between Two Proportions
- ~~For the Ratio of Two Variances (Normal Populations)~~ X

Comparing Means with Independent Samples

Two Sample Hypothesis Tests

Parameters of Interest

Difference in population means : $\mu_1 - \mu_2$

'' '' '' proportions: $p_1 - p_2$

Setting up hypotheses

Parameter of interest from population 1: θ_1

'' '' '' '' '' 2: θ_2

1/ Interested whether $\theta_1 > \theta_2$

$H_0: \theta_1 = \theta_2$ $\qquad$ $H_a: \theta_1 > \theta_2$

$H_0: \theta_1 - \theta_2 = 0$ $\qquad$ $H_a: \theta_1 - \theta_2 > 0$

2/ Interested whether $\theta_1 < \theta_2$

$H_0: \theta_1 = \theta_2$ $H_a: \theta_1 < \theta_2$

$H_0: \theta_1 - \theta_2 = 0$ $H_a: \theta_1 - \theta_2 < 0$

3/ Interested whether $\theta_1 \neq \theta_2$

$H_0: \theta_1 = \theta_2$ $H_a: \theta_1 \neq \theta_2$

$H_0: \theta_1 - \theta_2 = 0$ $H_a: \theta_1 - \theta_2 \neq 0$

Recall basic skeleton of a test statistic

$$\text{test stat} = \frac{(\text{A statistic}) - \left(\begin{array}{l}\text{hypothesized value}\\ \text{of parameter under } H_0\end{array}\right)}{\begin{array}{c}\text{Standard error}\\ \text{of statistic}\end{array}}$$

$\underline{\sigma \text{ known}}$

$$z^* = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

$\underline{\sigma \text{ unknown}}$

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

$\underline{\text{proportions}}$

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

Hypothes Test on a Difference on Means

$$\underbrace{\text{Difference on Means}}_{\mu_1 - \mu_2}$$

1) when σ_1 and σ_2 both known

$H_0: \mu_1 - \mu_2 = 0$ $\quad$ $H_a: \mu_1 - \mu_2 > 0$

$H_0: \mu_1 - \mu_2 = 0$ $\quad$ $H_a: \mu_1 - \mu_2 < 0$

$H_0: \mu_1 - \mu_2 = 0$ $\quad$ $H_a: \mu_1 - \mu_2 \neq 0$

Test stat

$$z^* = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

reference distribution: Standard normal

very rare to use since it is unlikely to know both σ_1 and σ_2

2/ when σ_1 and σ_2 both unknown

Case I : $\sigma_1 = \sigma_2$ assume equal st. devs

$(\sigma_1^2 = \sigma_2^2$, " " vars)

$H_0: \mu_1 - \mu_2 = 0$ $\quad H_a: \mu_1 - \mu_2 > 0$

$H_0: \mu_1 - \mu_2 = 0$ $\quad H_a: \mu_1 - \mu_2 < 0$

$H_0: \mu_1 - \mu_2 = 0$ $\quad H_a: \mu_1 - \mu_2 \neq 0$

test stat

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - 0}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{(\bar{X}_1 - \bar{X}_2) - 0}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

where

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

Reference distribution : t - distribution at $n_1 + n_2 - 2$ df

Comparing Means of Independent Samples (Normal Population Assumptions)

We should check the assumption that the underlying populations of individual responses are each Normally distributed. Nearly Normal Condition:

- We must check this for both groups; a violation for either one violates the condition.
- The Normality assumption matters most when sample sizes are very small.
- For $n < 10$ in either group, this method should not be used if the histogram or Normality plots show clear skewness.
- For n's of 10 or so, a moderately skewed histogram is okay. But, for strongly skewed data or data containing outliers this method should be avoided.
- For larger samples $n \geq 20$, data skewness is less of an issue - but, we still need to check if there are any outliers in the data, extreme skewness, and multiple modes.

Comparing Two Populations Means: Independent Sampling (Equal Variances Assumed)

Consider two independent populations with unknown means μ_1 and μ_2, and unknown standard deviations σ_1 and σ_2 ($\sigma_1 = \sigma_2$), respectively. We can make an inference about their mean difference $\mu_1 - \mu_2$ by using the difference between their point estimates (sample means): $\bar{Y}_1 - \bar{Y}_2$. When the assumptions and conditions are met,

$$\frac{(\bar{Y}_1 - \bar{Y}_2) - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{(\bar{Y}_1 - \bar{Y}_2) - (\mu_1 - \mu_2)}{\sqrt{S_p^2\left[\frac{1}{n_1} + \frac{1}{n_2}\right]}},$$

can be modelled by a $t(\nu)$ distribution; where $\nu = n_1 + n_2 - 2$ and $S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$.

Comparing Two Populations Means: Independent Sampling (Equal Variances Assumed) (cont.)

Conditions Required for Valid Inference about $\mu_1 - \mu_2$:

1. The two samples are randomly selected in an independent manner from the two target populations.
2. Both sampled populations have distributions that are approximately Normal.
3. The population variances are equal (e.g., $(\sigma_1 = \sigma_2)$).

Small-Sample Confidence interval for $\mu_1 - \mu_2$ (with equal variances)

Parameter : $\mu_1 - \mu_2$.

Confidence interval ($\nu = \text{df}$) :

$$(\bar{Y}_1 - \bar{Y}_2) \pm t_{\alpha/2}(\nu) S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}},$$

where $\nu = n_1 + n_2 - 2$ and $S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$
(requires that Normal samples are independent and the assumption that $\sigma_1^2 = \sigma_2^2$). The critical value $t_{\alpha/2}(\nu)$ depends on particular confidence level and the number of degrees of freedom.

Example

Comparing Two Population Means Managerial Success Indexes for Two Groups (With Equal Variances Assumed)
Behavioural researchers have developed an index designed to measure managerial success. The index (measured on a 100-point scale) is based on the manager's length of time in the organization and their level within the term; the higher the index, the more successful the manager. Suppose a researcher wants to compare the average index for the two groups of managers at a large manufacturing plant. Managers in group 1 engage in high volume of interactions with people outside the managers' work unit (such interaction include phone and face-to-face meetings with customers and suppliers, outside meetings, and public relation work). Managers in group 2 rarely interact with people outside their work unit. Independent random samples of 12 and 15 managers are selected from groups 1 and 2, respectively, and success index of each is recorded.

Example

Comparing Two Population Means Managerial Success Indexes for Two Group (With Equal Variances Assumed)
Note: The response variable is "Managerial Success Indexes".

- Managerial success indexes is a continuous quantitative variable, measured on 100-point scale.

The explanatory variable is "Type of group".

- Type of group (Group 1: Interaction with outsiders, Group 2: Fewer interactions) is a nominal categorical variable.

R Code

```
# Importing data file into R;

success=read.csv(file="success.csv",header=TRUE);

# Getting names of variables;

names(success);

# Seeing first few observations;

head(success);

# Attaching data file;
attach(success);
```

R Code

```
## [1] "Success_Index" "Group"
##   Success_Index Group
## 1            65     1
## 2            66     1
## 3            58     1
## 4            70     1
## 5            78     1
## 6            53     1
```

R Code (Descriptive Statistics)

```
# loading library mosaic;

library(mosaic);

favstats(Success_Index~Group);
```

Note. Group 1 = "interaction with outsiders" and Group 2 = "fewer interaction".

R Code (Descriptive Statistics)

$\bar{x}_1$ s_1 n_1

```
##    .group min    Q1 median    Q3 max     mean       sd  n mi
## 1       1  53 62.25   65.5 69.25  78 65.33333 6.610368 12
## 2       2  34 42.50   50.0 54.50  68 49.46667 9.334014 15
```

Note. Group 1 = “interaction with outsiders” and Group 2 = “fewer interactions”.

$\bar{x}_2$ s_2 n_2

R Code (Descriptive Statistics)

```
# WITHOUT library ;

summary(Success_Index[Group==1]);
length(Success_Index[Group==1]);
sd(Success_Index[Group==1]);

summary(Success_Index[Group==2]);
length(Success_Index[Group==2]);
sd(Success_Index[Group==2]);
```

R Code (Descriptive Statistics)

```
##     Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    53.00   62.25   65.50   65.33   69.25   78.00
## [1] 12
## [1] 6.610368
##     Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    34.00   42.50   50.00   49.47   54.50   68.00
## [1] 15
## [1] 9.334014
```

Nearly Normal Condition (Group 1: "interaction with outsiders"):

```
stem(Success_Index[Group==1]);
```

```
##
##   The decimal point is 1 digit(s) to the right of the |
##
##   5 | 38
##   6 | 0335689
##   7 | 018
```

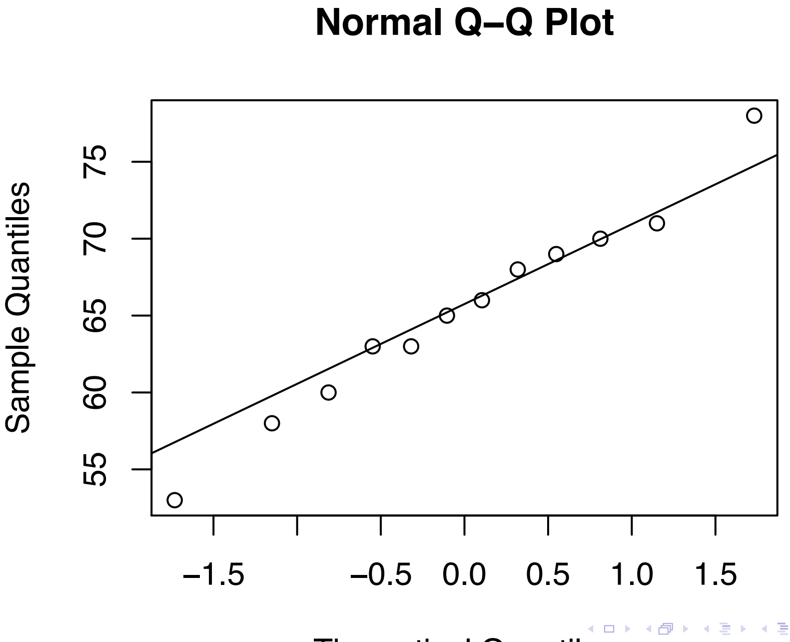
Nearly Normal Condition (Group 2: “fewer interactions”):

```
stem(Success_Index[Group==2]);
```

```
##
##   The decimal point is 1 digit(s) to the right of the |
##
##   3 | 46
##   4 | 22368
##   5 | 023367
##   6 | 28
```

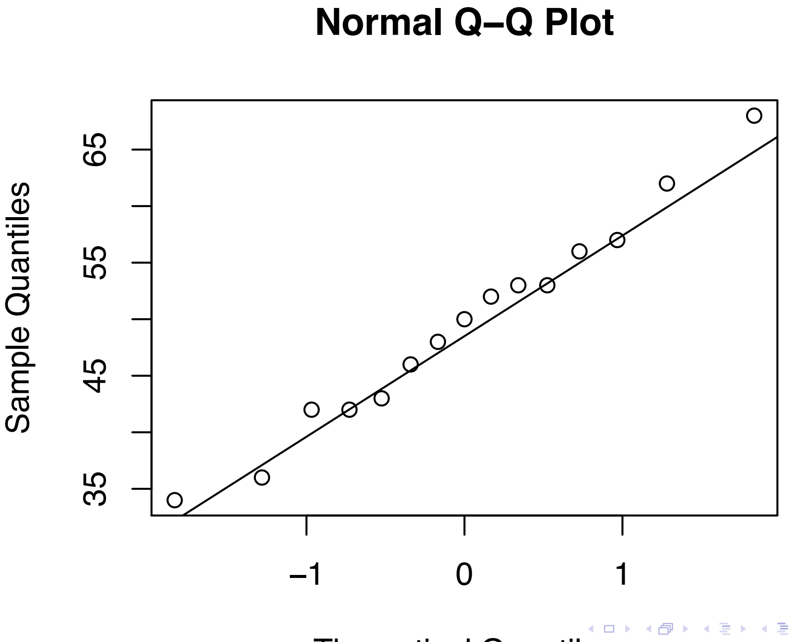
Nearly Normal Condition (Group 1: “interaction with outsiders”):

```
qqnorm(Success_Index[Group==1]);
qqline(Success_Index[Group==1]);
```


Normal Q–Q Plot
Sample Quantiles
55
60
65
70
75
–1.5
–0.5
0.0
0.5
1.0
1.5

Nearly Normal Condition (Group 2: “fewer interactions”):

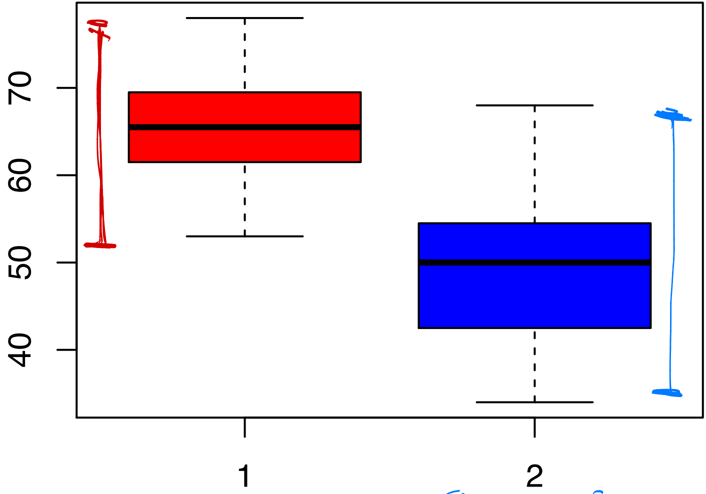
```
qqnorm(Success_Index[Group==2]);
qqline(Success_Index[Group==2]);
```


Normal Q–Q Plot
Sample Quantiles
35
45
55
65
–1
0
1

Nearly Normal Condition:

```
boxplot(Success_Index~Group,col=c("red","blue") )
```

Spread in boxplot 1 appears similar to spread in boxplot 2 (reasonable to assume equal vars)



approx symm

approx symm

Boxplot with ggplot2

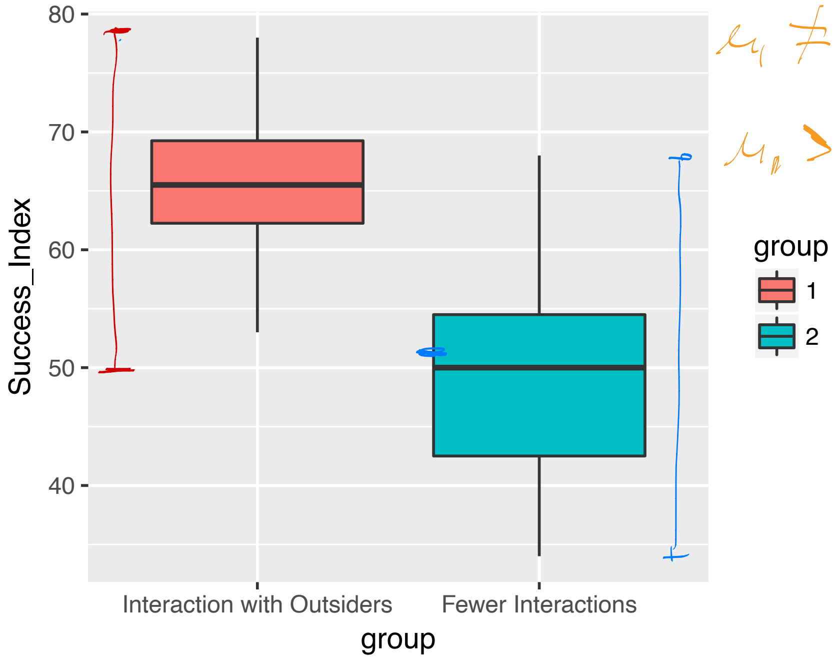
```
# loading library;
library(ggplot2);

# converting a numeric variable into factor (categorical data,
group<-factor(Group);

# bp: just a name (not code) to store boxplots;
bp<-ggplot(success,
aes(x=group,y=Success_Index, fill = group) );

our.labs=c("Interaction with Outsiders","Fewer Interactions");

bp +
geom_boxplot( )+
scale_x_discrete(labels = our.labs);
```



$\mu_1 \neq \mu_2$

$\mu_1 > \mu_2$ ✓

Example (Slide 18)

(select $\alpha = 0.05$)

	Group 1 (High interaction)	Group 2 (Low interaction)
	$n_1 = 12$	$n_2 = 15$
R output (slides 13, 15)	$\bar{X}_1 = 65.33$	$\bar{X}_2 = 49.467$
	$s_1 = 6.61$	$S_2 = 9.33$

Based on boxplots (Slides 28, 29) we decided to assume equal variances ($\sigma_1^2 = \sigma_2^2$)

(Reasonable → observed approx equal spread)

Asked to test for $\overbrace{\text{any difference}}^{\neq}$ in means (Slide 32)

$H_0: \mu_1 = \mu_2$ $\qquad$ $H_a: \mu_1 \neq \mu_2$

$H_0: \mu_1 - \mu_2 = 0$ $\qquad$ $H_a: \mu_1 - \mu_2 \neq 0$

Since we assumed $\sigma_1^2 = \sigma_2^2$, test stat

$$t^* = \frac{(\bar{X}_1 - \bar{X}_2) - 0}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

pooled variance

$$S_p^2 = \frac{(n_1 - 1) s_1^2 + (n_2 - 1) S_2^2}{n_1 + n_2 - 2}$$

$$= \frac{(12-1)\cdot 6.61^2 + (15-1)\cdot 9.33^2}{12+15-2}$$

$$= 67.97$$

pooled sample st. dev

$$S_p = \sqrt{S_p^2} = 8.244$$

test stat

$$t^* = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{65.33 - 49.67}{8.244\sqrt{\frac{1}{12} + \frac{1}{15}}}$$

$$= 4.97$$

reference distribution

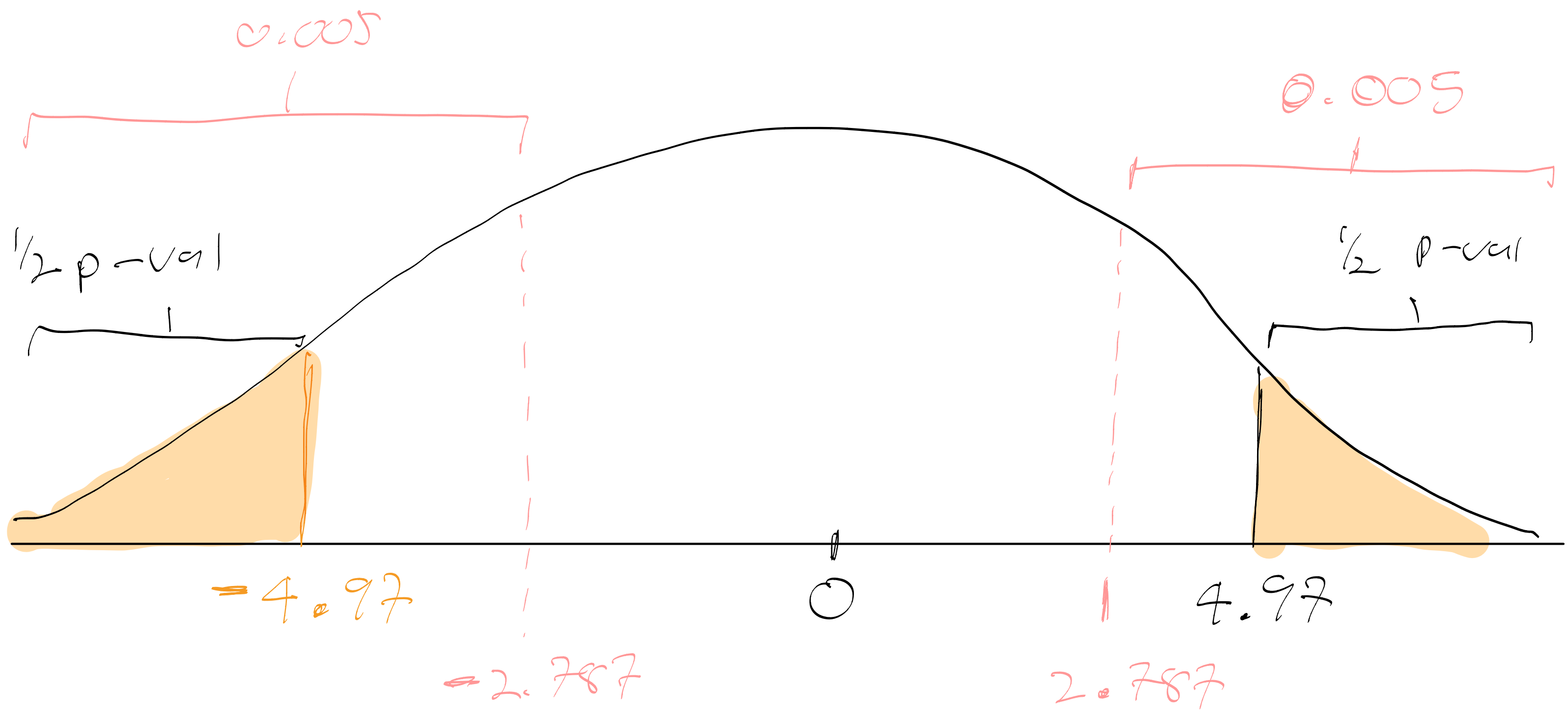
t distrib at $df = n_1 + n_2 - 2$

$$= 12 + 15 - 2$$

$$= 25$$

t distrib at 25 df

$H_a: \mu_1 - \mu_2 \neq$



one tail! $\frac{1}{2}$ p-val < 0.005

both tails p-val $< 0.01 < 0.05$ (α)

There is sufficient evidence to reject the null hypothesis that the mean index scores of the 2 groups is equal and conclude there is a difference.

Independent Group Assumption: The success index in group 1 is unrelated to success index in group 2. **Randomization Condition:** The 27 managers were randomly and independently selected (12 for group 1, and 15 for group 2).

Nearly Normal Condition: The two boxplots of success indexes do not show skewness; the two stemplots/histograms of success indexes are unimodal, fairly symmetric and approx. bell-shaped. Q-Q plots also suggest Normality assumption is reasonable.

Equal variances Assumptions: The two boxplots of success indexes appear to have the same spread; thus, the samples appear to have come from populations with approximately same variance.

Since the conditions are satisfied, it is appropriate to construct t CI with $df = 12 + 15 - 2 = 25$.

Example (cont.)

From the data, the following statistics were calculated:

$n_1 = 12$

$n_2 = 15$

$\bar{x}_1 = 65.33$

$\bar{x}_2 = 49.47$

$s_1^2 = 6.61^2$

$s_2^2 = 9.33^2$

Example (cont.)

The pooled variance estimator is

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \frac{(12 - 1)6.61^2 + (15 - 1)9.33^2}{12 + 15 - 2} = 67.97$$

Example (cont.)

The number of degrees of freedom of the test statistic is

$$\nu = n_1 + n_2 - 2 = 12 + 15 - 2 = 25$$

Example (cont.)

H_a: $>$ $<$ $\neq$

Two-sample t-test (Student's t-test) for the Difference between means $\mu_1 - \mu_2$.

Significance test about $\mu_1 - \mu_2$ (with Equal Population Variances Assumed).

Is there evidence to suggest that mean success index differ between the two groups?

$\neq$

Solution

1. State hypotheses.
$H_0 : \mu_1 = \mu_2$ or $H_0 : \mu_1 - \mu_2 = 0$
vs
$H_a : \mu_1 \neq \mu_2$ or $H_a : \mu_1 - \mu_2 \neq 0$,
where μ_1 is the mean success index for group 1 and μ_2 is the mean success index for group 2.

Solution

2. Test statistic.

$t^* = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}} = 4.97$

$(\bar{x}_1 = 65.33,\ \bar{x}_2 = 49.47\ ,\ s_p^2 = 67.97,\ n_1 = 12$ and $n_2 = 15)$

Solution

3. P-value.
Using Table, we have $df = n_1 + n_2 - 2 = 12 + 15 - 2 = 25$, and P-value $< 2(0.005)$.
Using R,

```
# one way;
2*(1 -pt(4.97, df=25));

## [1] 4.027976e-05

#another way;
2*pt(4.97,df=25,lower.tail=FALSE);

## [1] 4.027976e-05
```

Solution

4. Conclusion.

Since P-value is very small, we have strong evidence to indicate that there is a difference in mean success index between group 1 and group 2.
Note. As a follow-up, we could find a 95% CI for $\mu_1 - \mu_2$ to estimate this difference. Then, we could provide an estimate of how much higher the mean success index for group 1 is.

Hypothesis Test for $\mu_1 - \mu_2$

```
t.test(Success_Index ~ Group, var.equal=TRUE);

##
##  Two Sample t-test
##
## data:  Success_Index by Group
## t = 4.9675, df = 25, p-value = 4.055e-05
## alternative hypothesis: true difference in means is not equ
## 95 percent confidence interval:
##   9.288254 22.445079
## sample estimates:
## mean in group 1 mean in group 2
##        65.33333        49.46667
```

know we assumed $\sigma_1^2 = \sigma_2^2$

[Welch, then assumed $\sigma_1^2 \neq \sigma_2^2$]

$\neq$

$H_a: \mu_1 - \mu_2 \neq$

Case 2: when $\sigma_1 \neq \sigma_2$ unequal St. devs

$(\sigma_1^2 \neq \sigma_2^2$ " vars)

$H_0: \mu_1 - \mu_2 = 0$ $H_a: \mu_1 - \mu_2 > 0$

$H_0: \mu_1 - \mu_2 = 0$ $H_a: \mu_1 - \mu_2 < 0$

$H_0: \mu_1 - \mu_2 = 0$ $H_a: \mu_1 - \mu_2 \neq 0$

Test stat

$$t^* = \frac{(\bar{X}_1 - \bar{X}_2) - 0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

reference distribution:

t distribution where $df = \min(n_1 - 1, n_2 - 1)$

(by hand, software uses a more complicated approximation)

Small-Sample Confidence interval for $\mu_1 - \mu_2$ (unequal variances)

Draw an SRS of size n_1 from a Normal population with unknown mean μ_1, and draw and independent SRS of size n_2 from another Normal population with unknown mean μ_2. A confidence interval for $\mu_1 - \mu_2$ is given by

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

Here t^* is the critical value for the $t(k)$ density curve with area C between $-t^*$ and t^*. The degrees of freedom k are equal to the smaller of $n_1 - 1$ and $n_2 - 1$.

Degrees of freedom (Option 1)

Option 1. With software, use the statistic t with accurate critical values from the approximating t distribution.
The distribution of the two-sample t statistic is very close to the t distribution with degrees of freedom df given by

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

This approximation is accurate when both sample sizes n_1 and n_2 are 5 or larger.

Degrees of freedom (Option2)

Option 2. Without software, use the statistic t with critical values from the t distribution with degrees of freedom equal to the smaller of $n_1 - 1$ and $n_2 - 1$. These procedures are always conservative for any two Normal populations.

To test the hypothesis $H_0 : \mu_1 = \mu_2$, calculate the two-sample t statistic

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

and use P-values or critical values for the $t(k)$ distribution.

Degrees of freedom (Option 1)

Option 1. With software, use the statistic t with accurate critical values from the approximating t distribution.
The distribution of the two-sample t statistic is very close to the t distribution with degrees of freedom df given by

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

This approximation is accurate when both sample sizes n_1 and n_2 are 5 or larger.

Degrees of freedom (Option2)

Option 2. Without software, use the statistic t with critical values from the t distribution with degrees of freedom equal to the smaller of $n_1 - 1$ and $n_2 - 1$. These procedures are always conservative for any two Normal populations.

Logging in the rain forest

"Conservationists have despaired over destruction of tropical rain forest by logging, clearing, and burning". These words begin a report on a statistical study of the effects of logging in Borneo. Here are data on the number of tree species in 12 unlogged forest plots and 9 similar plots logged 8 years earlier:

Unlogged: 22 18 22 20 15 21 13 13 19 13 19 15

Logged : 17 4 18 14 18 15 15 10 12

Does logging significantly reduce the mean number of species in a plot after 8 years? State the hypotheses and do a t test. Is the result significant at the 5% level?

Solution

Does logging significantly reduce the mean number of species in a plot after 8 years?

1. State hypotheses. $H_0 : \mu_1 = \mu_2$ vs $H_a : \mu_1 > \mu_2$, where μ_1 is the mean number of species in unlogged plots and μ_2 is the mean number of species in plots logged 8 years earlier.
2. Test statistic. $t^* = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = 2.1140$ ($\bar{x}_1 = 17.5$, $\bar{x}_2 = 13.6666$, $s_1 = 3.5290$, $s_2 = 4.5$, $n_1 = 12$ and $n_2 = 9$)
3. P-value. Using Table, we have $df = 8$, and $0.025 <$ P-value < 0.05.
4. Conclusion. Since P-value < 0.05, we reject H_0. There is strong evidence that the mean number of species in unlogged plots is greater than that for logged plots 8 year after logging.

Example

$H_a: \mu_{trt} > \mu_{control}$

A company that sells educational materials reports statistical studies to convince customers that its materials improve learning. One new product supplies "directed reading activities" for classroom use. These activities should improve the reading ability of elementary school pupils.
A consultant arranges for a third-grade class of 21 students to take part in these activities for an eight-week period. A control classroom of 23 third-graders follows the same curriculum without the activities. At the end of the eight weeks, all students are given a Degree of Reading Power (DRP) test, which measures the aspects of reading ability that the treatment is designed to improve. The data appear in the following table.

Data

Treatment				Control			
24	61	59	46	42	33	46	37
43	44	52	43	43	41	10	42
58	67	62	57	55	19	17	55
71	49	54		26	54	60	28
43	53	57		62	20	53	48
49	56	33		37	85	42	

Because we hope to show that the treatment (Group 1) is better than the control (Group 2), the hypotheses are:

$H_0 : \mu_1 = \mu_2$

vs

$H_a : \mu_1 > \mu_2$

```r
# Step 1. Entering data;

treatment=c(24, 61, 59, 46, 43, 44, 52, 43, 58, 67,
62, 57, 71, 49, 54, 43, 53, 57, 49, 56, 33);

control=c(42, 33, 46, 37, 43, 41, 10, 42, 55, 19, 17,
55, 26, 54, 60, 28, 62, 20, 53, 48, 37, 85, 42);
```

Checking the assumptions

Nearly Normal Condition (treatment):

```
# Making stemplot;

stem(treatment);
```

Checking the assumptions

Nearly Normal Condition (treatment):

```
##
##   The decimal point is 1 digit(s) to the right of the |
##
##   2 | 4
##   3 | 3
##   4 | 3334699
##   5 | 23467789
##   6 | 127
##   7 | 1
```

Checking the assumptions

Nearly Normal Condition (control):

```r
# Making stemplot;

stem(control);
```

Checking the assumptions

Nearly Normal Condition (control):

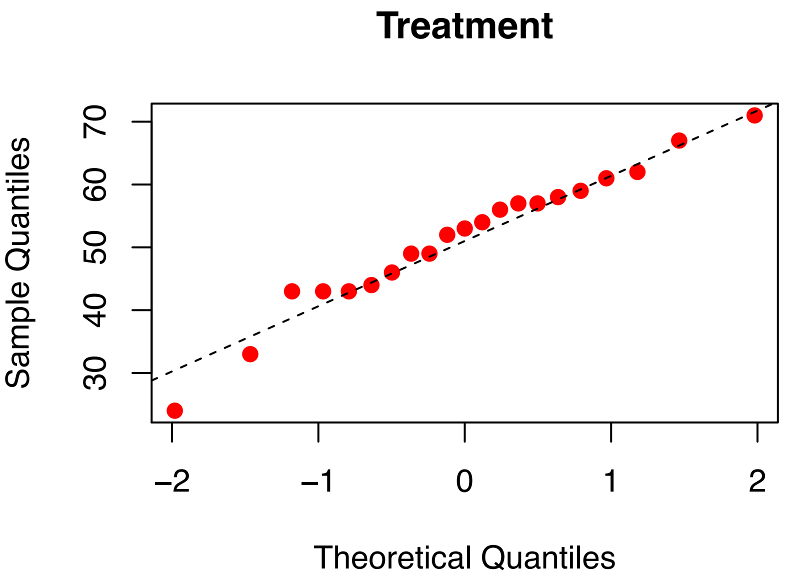
```
##
##   The decimal point is 1 digit(s) to the right of the |
##
##   0 | 079
##   2 | 068377
##   4 | 12223683455
##   6 | 02
##   8 | 5
```

Checking the assumptions

Nearly Normal Condition (treatment):

```
# Making Q-Q plot;
qqnorm(treatment,pch=19,col="red",main="Treatment");
qqline(treatment,lty=2);
```

Nearly Normal Condition (treatment):

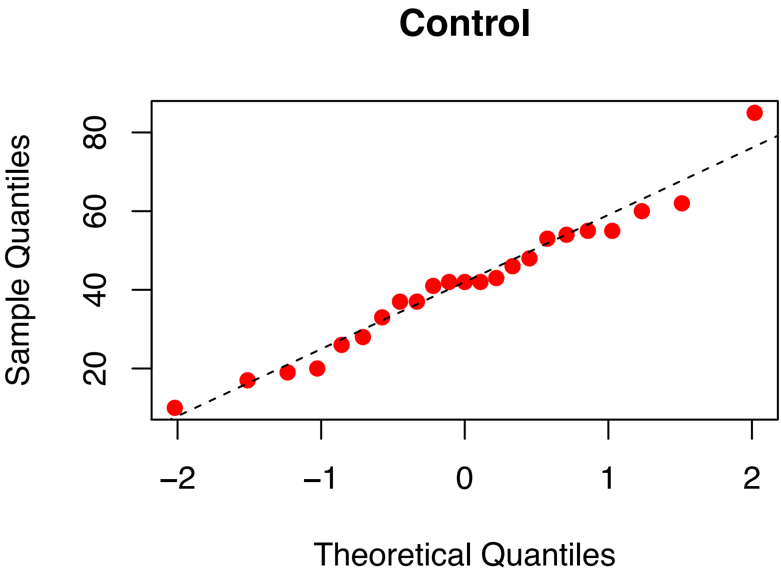


Checking the assumptions

Nearly Normal Condition (control):

```
# Making Q-Q plot;
qqnorm(control,pch=19,col="red",main="Control");
qqline(control,lty=2);
```

Nearly Normal Condition (control):



Stemplots suggest that there is a mild outlier in the control group but no deviation from Normality serious enough to prevent us from using t procedures. Normal Q-Q plots for both groups confirm that both are roughly Normal. The summary statistics are:

```
summary(treatment);

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   24.00   44.00   53.00   51.48   58.00   71.00

summary(control);

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   10.00   30.50   42.00   41.52   53.50   85.00
```

```r
# Step 1. Entering data;

treatment=c(24, 61, 59, 46, 43, 44, 52, 43, 58, 67,
62, 57, 71, 49, 54, 43, 53, 57, 49, 56, 33);

control=c(42, 33, 46, 37, 43, 41, 10, 42, 55, 19, 17,
55, 26, 54, 60, 28, 62, 20, 53, 48, 37, 85, 42);

# Step 2. Hypothesis Test;

t.test(treatment,control,alternative="greater");
```

assumned $\sigma_1 \neq \sigma_2$

```
##
##  Welch Two Sample t-test
##
## data:  x and control
## t = 2.3109, df = 37.855, p-value = 0.01319
## alternative hypothesis: true difference in means is greater
## 95 percent confidence interval:
##  2.691293      Inf
## sample estimates:
## mean of x mean of y
##  51.47619  41.52174
```

$H_a: \mu_1 - \mu_2 >$

HT (using table)

Summary statistics

```
round(mean(treatment),2);

## [1] 51.48

round(sd(treatment),2);

## [1] 11.01

round(mean(control),2);

## [1] 41.52

round(sd(control),2);

## [1] 17.15
```

Example (slide 46)

(select $\alpha = 0.05$)

	① Trt	② Control
	$n_1 = 21$	$n_2 = 23$
R output (slide 61)	$\bar{X}_1 = 51.48$	$\bar{X}_2 = 41.52$
	$s_1 = 11.01$	$s_2 = 17.15$

Asked to test whether treatment improves scores

$H_0: \mu_1 = \mu_2$ $\qquad H_a: \mu_1 > \mu_2$

$H_0: \mu_1 - \mu_2 = 0$ $\qquad H_a: \mu_1 - \mu_2 > 0$

We are considering $\sigma_1 \neq \sigma_2$ [later we'll learn a technique to determine this]

Test st

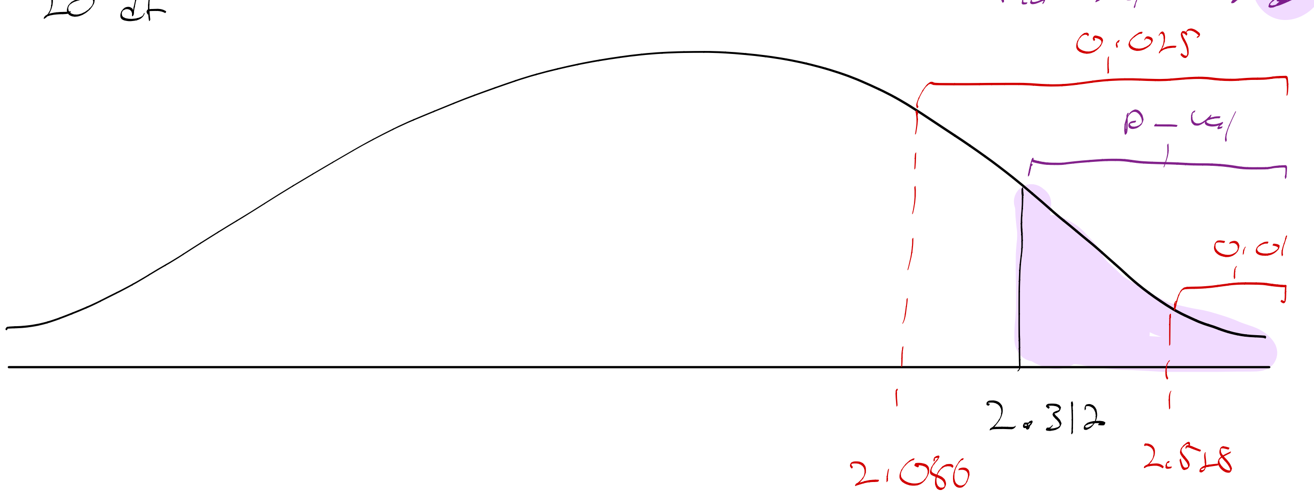
$$t^* = \frac{(\bar{X}_1 - \bar{X}_2) - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{51.48 - 41.52}{\sqrt{\frac{11.01^2}{21} + \frac{17.15^2}{23}}} = 2.312$$

Reference distrib

t dist at $\min(\underset{21-1}{n_1 - 1}, \underset{23-1}{n_2 - 1}) = 20$ df

t at 20 df

$H_a: \mu_1 - \mu_2 >$



$0.01 < \underbrace{p\text{-val} < 0.025 < \underset{(\alpha)}{0.05}}_{p\text{-val} < \alpha}$

Conclusion: Exercise (reject H_0, write in correct context

Test statistic

2. Test statistic.

$t^* = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = 2.31$ ($\bar{x}_1 = 51.48$, $\bar{x}_2 = 41.52$, $s_1 = 11.01$, $s_2 = 17.15$, $n_1 = 21$ and $n_2 = 23$)

P-value

The conservative approach uses the $t(20)$ distribution. The P-value for the one-sided test is

$$\text{P-value} = P(T \geq 2.31)$$

Comparing $t = 2.31$ with the entries in Table 5 for 20 degrees of freedom, we see that

$$0.01 < \text{P-value} < 0.025.$$

Conclusion

Since our P-value is "small", we reject the null hypothesis (Note that we would reject H_0 at the 2.5% significance level). The data strongly suggest that directed reading activity improves the DRP score.

Additional info about the DRP study

The design of the DRP study is not ideal. Random assignment of students was not possible in a school environment, so existing third-grade classes were used. The effect of the reading programs is therefore confounded with any other differences between the two classes. The classes were chosen to be as similar as possible in variables such as the social and economic status of the students. Pretesting showed that the two classes were on the average quite similar in reading ability at the beginning of the experiment. To avoid the effect of two different teachers, the same teacher taught reading in both classes during the eight-week period of the experiment. We can therefore be somewhat confident that our two-sample procedure is detecting the effect of the treatment and not some other difference between the classes.

Review

2 Sample Hypothesis Tests on Difference of Means $(\mu_1 - \mu_2)$

$H_0: \mu_1 - \mu_2 = 0$

$H_a: \mu_1 - \mu_2 > 0$, $\mu_1 - \mu_2 < 0$ — one-sided (one-tailed)

$\mu_1 - \mu_2 \neq 0$ — two-sided (two-tailed)

1/ When σ_1 and σ_2 are both known
(very rare, unrealistic situation)

test stat
$$z^* = \frac{(\bar{X}_1 - \bar{X}_2) - 0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

reference dist: standard normal

2/ When σ_1 and σ_2 are both unknown

Case 1: Equal variances $\sigma_1^2 = \sigma_2^2$ — pooled method
(" st. devs $\sigma_1 = \sigma_2$)

test stat
$$t^* = \frac{(\bar{X}_1 - \bar{X}_2) - 0}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{(\bar{X}_1 - \bar{X}_2) - 0}{\sqrt{S_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

where
$$S_p^2 = \frac{(n_1 - 1)\cdot S_1^2 + (n_2 - 1) S_2^2}{n_1 + n_2 - 2}$$ (pooled variance)

$S_p = +\sqrt{S_p^2}$ (pooled st dev)

reference dist: t-distribution of $n_1 + n_2 - 2$ df

Case 2: Unequal variances $\sigma_1^2 \neq \sigma_2^2$ $\boxed{\text{WELCH}}$

(" st. dev's $\sigma_1 \neq \sigma_2$)

test stat

$$t^* = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

reference distribution:

t-distribution at $\min(n_1 - 1, n_2 - 1)$ df (by hand)

[R uses a more sophisticated approximation for df for this case]

Why are we doing this?

To conduct tests on parameters and to quantify the degree of certainty with probability.

The smaller the p-value, the stronger the evidence against H_0 (provided assumptions are met)

Suppose you had the following limited information

1/ H_0 rejected at 5% level. p-value < 0.05

A test was statistically significant at the 5% level. Can we also conclude this test is significant at the 1% level?

p-value < 0.01 ?

We can not make this conclusion for certain

$$(\text{p-value} < 0.05) \not\Rightarrow (\text{p-value} < 0.01)$$

2/ p-value < 0.01

A test was statistically significant at the 1% level. Can we also conclude this test is significant at the 5% level?

p-value < 0.05 ?

Yes!

$$\text{p-value} < 0.01 < 0.05$$

The Fold Rule (pdf slide 66)

A rule which can be used to quickly determine whether population variances are equal or unequal using sample variances

[Note: only a rule]

↓

If $\dfrac{\max(s_1, s_2)}{\min(s_1, s_2)} < \sqrt{2}$ then we can consider $\sigma_1^2 = \sigma_2^2$

$\dfrac{\max(s_1^2, s_2^2)}{\min(s_1^2, s_2^2)} < 2$ then we can consider $\sigma_1^2 = \sigma_2^2$

Note! rather crude technique

Hypothesis tests on 2 variances exist

$H_0: \sigma_1^2 = \sigma_2^2$ $\quad$ $H_a: \sigma_1^2 \neq \sigma_2^2$

Involves χ^2 test stat

Comparing Means with Paired Samples

2 dependent measurements on each unit.

use the differences

Hypothesis Test on Paired Data

Unit	Measurement 1	Measurement 2	Difference (Measurement 2 − Measurement 1)
1	x_{11}	x_{12}	$d_1 = x_{12} - x_{11}$
2	x_{21}	x_{22}	$d_2 = x_{22} - x_{21}$
⋮	⋮	⋮	⋮
n	x_{n1}	x_{n2}	$d_n = x_{n2} - x_{n1}$

$\bar{x}_d$ mean difference

s_d st. dev of differences

Hyp. Test

$H_0: \mu_D = 0 \quad (\mu_2 = \mu_1)$ $\qquad$ $H_a: \mu_D > 0 \quad (\mu_2 > \mu_1)$

$H_0: \mu_D = 0$ $\qquad$ $H_a: \mu_D < 0 \quad (\mu_2 < \mu_1)$

$H_0: \mu_D = 0$ $\qquad$ $H_a: \mu_D \neq 0 \quad (\mu_2 \neq \mu_1)$

Test statistic

$$t^* = \frac{\bar{x}_d - 0}{s_d/\sqrt{n}}$$

reference distribution: t distribution of n−1 df

[Very similar to one-sample t-test.
However it is important to recognize
we are working with 2 dependent measurements]

Inference Procedures for the Population Mean μ

One-Sample

Inference Procedure for μ when σ is known

Confidence interval for μ:

$$\bar{x} \pm Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

To test $H_0: \mu = \mu_0$, test statistic is:

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

Inference Procedure for μ when σ is <u>not</u> known

For paired diff

Confidence interval for μ:

$$\bar{x}_d \pm t_{(\alpha/2,\, n-1)}\frac{s_d}{\sqrt{n}}$$

To test $H_0: \mu = \mu_0$, test statistic is:

$$t^* = \frac{\bar{x}_d - \mu_0}{s_d/\sqrt{n}}$$

Two-Sample

Inference Procedure for Difference between μ_1 and μ_2 when σ_1 and σ_2 are known

Confidence interval for $\mu_1 - \mu_2$:

$$(\bar{x_1} - \bar{x_2}) \pm Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

To test $H_0: \mu_1 - \mu_2 = \mu_D$, test statistic is:

$$Z^* = \frac{(\bar{x_1} - \bar{x_2}) - \mu_D}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Inference Procedure for Difference between μ_1 and μ_2 when σ_1 and σ_2 are <u>not</u> known

<u>Case 1: Unequal Variances</u>

Confidence interval for $\mu_1 - \mu_2$:

$$(\bar{x_1} - \bar{x_2}) \pm t_{(\alpha/2,\ d)}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

To test $H_0: \mu_1 - \mu_2 = \mu_D$, test statistic is:

$$t^* = \frac{(\bar{x_1} - \bar{x_2}) - \mu_D}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Comparing Means with Paired Samples

- When observations in sample 1 matches with an observation in sample 2.
- Observations in sample 1 are, usually, highly, correlated with observations in sample 2.
- The data are often called matched pairs.
- For each pair (the same cases), we form: Difference = observation in sample 2 - observation in sample 1.
- Thus, we have one single sample of differences scores.
- For example, in longitudinal studies: Pre- and post-survey of attitudes towards statistics (Same student is measured twice: Time 1 (pre) and Time 2 (post). We measure change in the attitudes: Post - Pre (for each student).
- Often these types of studies are called, *repeated measures* .

Paired Samples: Assumptions and Conditions

Paired Data Condition:

- The data must be quantitative and paired.

Independence Assumption:

- If the data are paired, the groups are not independent. For this methods, it is the differences that must be independent of each other.
- The pairs may be a random sample.
- In experimental design, the order of the two treatments may be randomly assigned, or the treatments may be randomly assigned to one member of each pair.
- In a before-and-after study, we may believe that the observed differences are representative sample of a population of interest. If there is any doubt, we need to include a control group to be able to draw conclusions.

Paired Samples: Assumptions and Conditions

Independence Assumption (cont.):

- If samples are bigger than 10% of the target population, we need to acknowledge this and note in our report. When we sample from a finite population, we should be careful not to sample more than 10% of that population. Sampling too large a fraction of the population calls the independence assumption into question.

Paired Samples: Assumptions and Conditions

Normal Population Assumption
We assume that the population of differences follows a Normal model. We need to check:
Nearly Normal Condition:

- This condition can be checked with a histogram or Boxplot of differences - but not of the individual groups.
- As with the case of the one-sample t-methods, robustness against departure form normality increases with sample size; in other words, the Normality assumptions matter less the more pairs we have to consider.

Note: When the conditions are met, we can model the sampling distribution of difference in sample means with a Student's t-model with n- 1 degrees of freedom.

Confidence Interval: Paired t-Interval

When the assumptions and conditions are met, the confidence interval for the mean of paired difference $\mu_1 - \mu_2$ is:
Point Estimate $\pm$ Margin of Error of the Point Estimate

$$\bar{X}_d \pm t^*_{df} SE(\bar{X}_d)$$

Where the standard error of the mean difference is estimated as

$$SE(\bar{X}_d) = \frac{s_d}{\sqrt{n}}$$

The critical value t^*_{df} depends on the particular confidence level and the number of df = n - 1, which is based on the number of pairs, n.

Matched pairs t procedures

To compare the responses to the two treatments in a matched pairs design, find the difference between the responses within each pair. Then apply the one-sample t procedures to these differences.
A matched pairs design compares just two treatments. Choose pairs of subjects that are as closely matched as possible. Assign one of the treatments to one of the subjects in a pair by tossing a coin or reading odd and even digits from a table of random digits (or by generating them with a computer). The other subject gets the remaining treatment. Sometimes each "pair" in a matched pairs design consists of just one subject, who gets both treatments one after the other.

Example

A manufacturer wanted to compare the wearing qualities of two different types of automobile tires, A and B. In the comparison, a tire of type A and one of type B were randomly assigned and mounted on the rear wheels of each of five automobiles. The automobiles were then operated for a specified number of miles, and the amount of wear was recorded for each tire. These measurements appear in a table below. Do the data provide sufficient evidence to indicate a difference in mean wear for tire types A and B? Test using $\alpha = 0.05$. Assume that variable wear is Normally distributed.

Auto	1	2	3	4	5
Tire A	10.6	9.8	12.3	9.7	8.8
Tire B	10.2	9.4	11.8	9.1	8.3

Solution

You can verify that the mean and standard deviation of the five **difference** measurements are $\bar{d} = 0.48$ and $s_d = 0.0837$.

Step 1. State Hypotheses. $H_0 : \mu_d = 0$ vs $H_a : \mu_d \neq 0$.
Step 2. Find test statistic. $t^* = \frac{\bar{d}-0}{s_d/\sqrt{n}} = \frac{0.48}{0.0837/\sqrt{5}} = 12.8$
Step 3. Compute P-value. Using Table, $P - value < 2(0.005)$
Step 4. Conclusion. Since $P - value < \alpha = 0.05$, we reject H_0. There is ample evidence of a difference in the mean amount of wear for tire types A and B.

Example

Find a 95% confidence interval for $(\mu_A - \mu_B) = \mu_d$ using the data from our previous example.

Solution

A 95% confidence interval for the difference between the mean wear is

$$\bar{d} \pm t^* \frac{s_d}{\sqrt{n}}$$

$$0.48 \pm (2.776)\frac{0.0837}{\sqrt{5}}$$

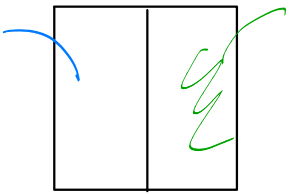
$$0.48 \pm 0.1039$$

Exercise

In an effort to determine whether a new type of fertilizer is more effective than the type currently in use, researchers took 12 two-acre plots of land scattered throughout the county. Each plot was divided into two equal-size subplots, one of which was treated with the new fertilizer. Wheat was planted, and the crop yields were measured.

Current

new



dependency is on the plot

Exercise

Plot	1	2	3	4	5	6	7	8	9	10	11	12
Current	56	45	68	72	61	69	57	55	60	72	75	66
New	60	49	66	73	59	67	61	60	58	75	72	68

Exercise

a. Can we conclude at the 5% significance level that the new fertilizer is more effective than the current one?

Example Slide 77-78 $\alpha = 0.05$

Measurements depend on plot $\longrightarrow$ paired t-test

plot	Current	New	Differences (New - Current)
1	56	60	60 - 56 = 4
2	45	49	49 - 45 = 4
3	68	66	66 - 68 = -2
⋮	⋮	⋮	⋮
12	66	68	68 - 66 = +2

$n = 12$

$\bar{X}_d = 1$

$S_d = 3.015$

Test whether new gives better yield

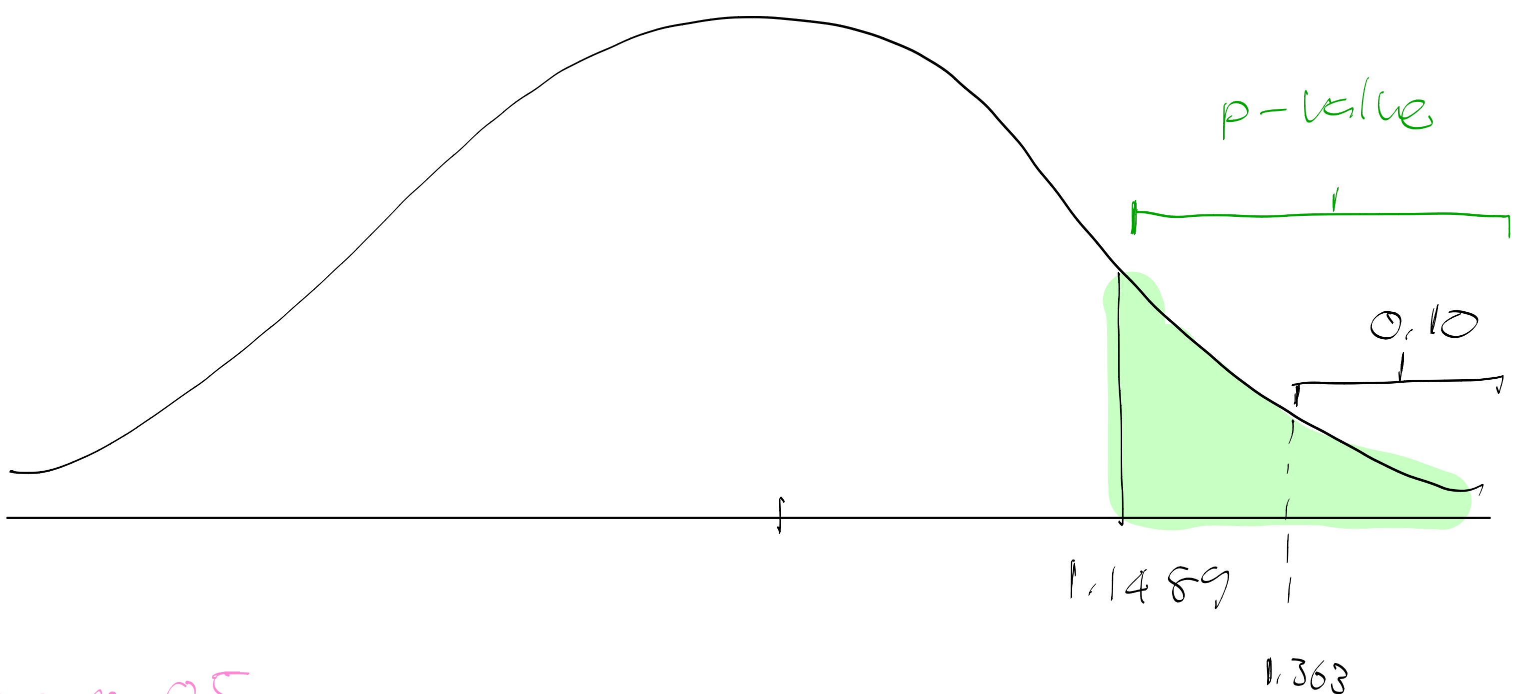
$H_0: \mu_d = 0$ $H_a: \mu_d > 0$

test stat

$$t^* = \frac{\bar{X}_d - 0}{S_d / \sqrt{n}} = \frac{1 - 0}{3.015/\sqrt{12}} = 1.1489$$

reference distribution: t distribution at $n-1 = 11$ df t at 11 df

$H_a: \mu_d >$



$\alpha = 0.05$

$$\text{p-value} > 0.10 > \underset{\alpha}{0.05}$$

Insufficient evidence to reject Ho.
Do not reject the null hypothesis that the 2 fertilizers give equal crop yields.

Solution a)

You can verify that the mean and standard deviation of the twelve **difference** measurements are $\bar{d}$ = new - current = 1 and $s_d = 3.0151$.

Step 1. State Hypotheses. $H_0 : \mu_d = 0$ vs $H_a : \mu_d > 0$.
Step 2. Find test statistic. $t^* = \frac{\bar{d}-0}{s_d/\sqrt{n}} = \frac{1}{3.0151/\sqrt{12}} = 1.1489$
Step 3. Compute P-value. Using Table (df=11), $0.10 < P - value < 0.15$.
Exact P-value = 0.1375, using R.
Step 4. Conclusion. Since $P - value > \alpha = 0.05$, we **can't** reject H_0.
There is not enough evidence to infer that the new fertilizer is better.

R Code

```
# Step 1. Entering data;

current=c(56, 45, 68, 72, 61, 69, 57, 55, 60, 72, 75, 66);

new=c(60, 49, 66, 73, 59, 67, 61, 60, 58, 75, 72, 68);

diff=new-current;

# Step 2. T test;

t.test(diff,alternative="greater");
```

R Code

```
##
##  One Sample t-test
##
## data:  x
## t = 1.1489, df = 11, p-value = 0.1375
## alternative hypothesis: true mean is greater than 0
## 95 percent confidence interval:
##  -0.5631171        Inf
## sample estimates:
## mean of x
##         1
```

R Code (Another way)

```
# Step 1. Entering data;

current=c(56, 45, 68, 72, 61, 69, 57, 55, 60, 72, 75, 66);

new=c(60, 49, 66, 73, 59, 67, 61, 60, 58, 75, 72, 68);

# Step 2. T test;

t.test(new,current,paired=T, alternative="greater");
```

R Code (Another way)

```
##
##  Paired t-test
##
## data:  x and current
## t = 1.1489, df = 11, p-value = 0.1375
## alternative hypothesis: true difference in means is greater
## 95 percent confidence interval:
##  -0.5631171        Inf
## sample estimates:
## mean of the differences
##                       1
```

than 0

Interpret output

Hyp test on paired data

$H_0: \mu_d = 0$ $H_a: \mu_d > 0$

test stat: $t = 1.1469$

reference dist: t at 11 df

p-value 0.1375

Example

Comparing 2016 and 2017 Voter Turnout % in OCED Countries
The "Better Life Index" program (BLI, 2017 and 2016) includes set of indicators regarding social protection and well-being in OECD countries. A quantitative variable named "Voter Turnout" is a sub-component of the component in this BLI program, which is defined as the ratio between the number of individuals that cast a ballot during an election (whether this vote is valid or not) to the population registered to vote. Thus, the unit of measurement is population percentage of voter turnout in OECD countries. As institutional features of voting systems vary a lot across countries and across types of elections, the indicator refers to the elections (parliamentary or presidential) that have attracted the largest number of voters in each country. OECD indicates that they obtained this information from International Institute for Democracy and Electoral Assistance (IDEA); Comparative Studies of Electoral System for inequality data (self-reported voter turnout).

R Code

```
# Importing data file into R;

voterT=read.csv(file="VoterTurnout.csv",header=TRUE);

# Getting names of variables;

names(voterT);

# Seeing first few observations;

head(voterT);

# Attaching data file;
attach(voterT);
```

R Code

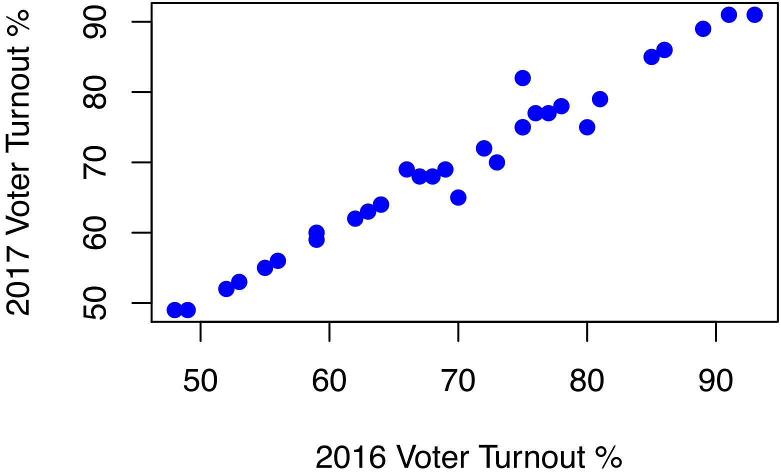
```
## [1] "Country"             "Voter_turnout_2017" "Voter_turnou
##          Country Voter_turnout_2017 Voter_turnout_2016
## 1      Australia                 91                 93
## 2        Austria                 75                 75
## 3        Belgium                 89                 89
## 4         Canada                 68                 68
## 5          Chile                 49                 49
## 6 Czech Republic                 59                 59
```

Association 2016 and 2017 Voter Turnout % in OECD Countries

```r
# Scatterplot of data;
plot(Voter_turnout_2016, Voter_turnout_2017,
xlab =" 2016 Voter Turnout %",
ylab="2017 Voter Turnout %", pch=19, col="blue");

# Sample Correlation, r;
cor(Voter_turnout_2016, Voter_turnout_2017);
```

Association 2016 and 2017 Voter Turnout % in OECD Countries



Association 2016 and 2017 Voter Turnout % in OECD Countries (Sample Correlation)

```
## [1] 0.9868955
```

Plot Interpretation: As percentage of voter turnout in 2016 increases, percentage of voter turnout in 2017 tend to increase.
Estimated Correlation Interpretation: If an OECD country's 2016 voter turnout is 1 standard deviation above the mean, its 2017 voter turnout percentage is approx. 0.987 standard deviation above the mean 2017 voter turnout %.

Working with Summary Statistics

Let μ_1 denote the population mean voter turnout percentages in 2017. Let $\bar{X}_1$ denote the sample mean voter turnout percentages in 2017, that estimates μ_1.
Let μ_2 denote the population mean voter turnout percentages in year 2016. Let $\bar{X}_2$ denote the sample mean voter turnout percentages in 2016, that estimates μ_2.
Let μ_d denote the population mean difference in voter turnout percentages ($\mu_d = \mu_1 - \mu_2$). Let $\bar{X}_d$ denote the sample mean of the difference in voter turnout percentages, that estimates μ_d.
Note: The mean of the estimated differences $\bar{X}_d$ equals the differences between the estimated means $\bar{X}_1 - \bar{X}_2$.

Summary statistics using R

```
# Obtaining summary statistics;

# loading library;
library(mosaic);
favstats(Voter_turnout_2017);

favstats(Voter_turnout_2016);

Diff_Voter_turnout = Voter_turnout_2017 - Voter_turnout_2016;
favstats(Diff_Voter_turnout);
```

Summary statistics using R

```
## min    Q1 median    Q3 max     mean        sd  n missing
##  49 62.25   69.5 77.75  91 70.14706 12.06832 34       0
## min    Q1 median    Q3 max     mean        sd  n missing
##  48 62.25     71 77.75  93 70.23529 12.24017 34       0
## min Q1 median Q3 max        mean       sd  n missing
##  -5  0      0  0   7 -0.08823529 1.975113 34       0
```

Summary statistics using R

```
# Obtaining summary statistics;

# WITHOUT loading library;

Diff_Voter_turnout = Voter_turnout_2017 - Voter_turnout_2016;
mean(Diff_Voter_turnout);
sd(Diff_Voter_turnout);
```

Summary statistics using R

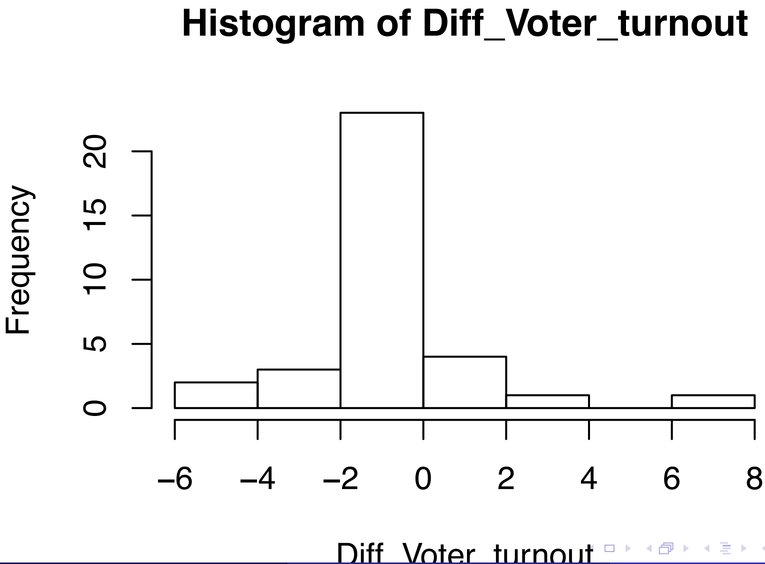
```
## [1] -0.08823529
## [1] 1.975113
```

Checking Assumptions and Conditions

```
Diff_Voter_turnout = Voter_turnout_2017 - Voter_turnout_2016;

# Making histogram;
hist(Diff_Voter_turnout);
```

Checking Assumptions and Conditions



Checking Assumptions and Conditions

Paired Data Condition:
The data are paired because we are interested in difference voter turnout percentages in 2016 and 2017 within each OECD country.
Independence/Randomization:
Each pair (2016 voter turnout %, 2017 voter turnout %) is independent of the others, so the differences are independent.
Nearly Normal Condition:
The histogram of the differences is approx. bell-shaped and symmetric.
The conditions are met; so, we can use a Student's t-model with n-1 degrees of freedom to construct confidence intervals.

Example of a Paired t-test (Two-Sided Test)

Inferences Comparing Means Using Paired Differences (Hypothesis Testing).
Is there evidence to suggest that percentages of voter turnouts in OECD countries are different in 2016 and 2017?

Step 1. State Hypotheses.
$H_0 : \mu_d = 0$ (no difference between mean voter turnout % in 2017 and 2016)
vs
$H_a : \mu_d > 0$ (there is a difference between voter turnout % in 2017 and 2016).

Step 2. Find test statistic.

$$t^* = \frac{\bar{d} - 0}{s_d/\sqrt{n}} = \frac{-0.09}{1.98/\sqrt{34}} = -0.26$$

Step 3. Compute P-value.
Using Table (df=33 $\approx$ 29),

$$P - value = 2 * P[T > |t^*|] = 2 * P[T > 0.26]$$

$P - value > 0.20$.
Using R,

```
2*pt(-0.26,df=33);

## [1] 0.7964792

2*( 1-pt(0.26,df=33) );

## [1] 0.7964792
```

Confirm results with R

```
t.test(Voter_turnout_2017, Voter_turnout_2016, paired=TRUE);
```

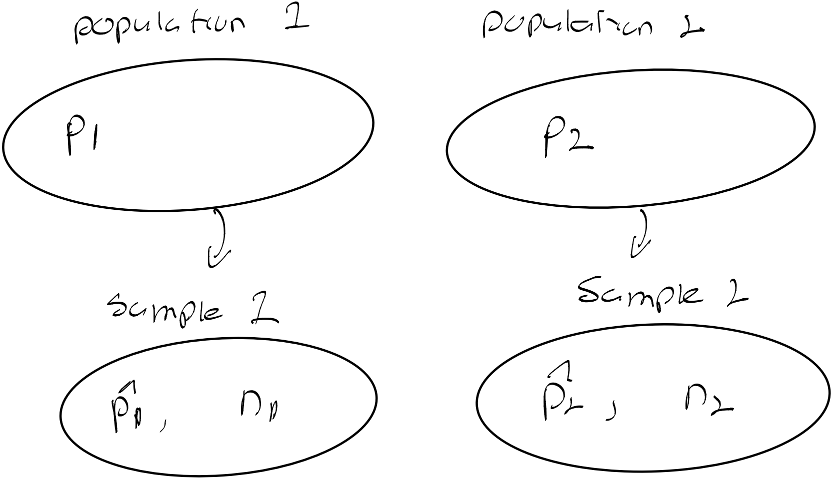
Confirm results with R

```
##
##  Paired t-test
##
## data:  x and Voter_turnout_2016
## t = -0.26049, df = 33, p-value = 0.7961
## alternative hypothesis: true difference in means is not equ
## 95 percent confidence interval:
##  -0.7773846  0.6009140
## sample estimates:
## mean of the differences
##             -0.08823529
```

Conclusion

$P - value > \alpha = 0.05$. We fail to reject $H_0 : \mu_d = 0$.
We have no evidence to conclude that there is a difference in 2017 and 2016' mean voter turnout percentages in OECD countries.ntries.

Comparing Proportions



Interested in $p_1 - p_2$ (difference in proportions)

Hypothesis Test on a Difference of Population Proportions $p_1 - p_2$

$H_0: p_1 - p_2 = 0$ $\quad$ $H_a: p_1 - p_2 > 0$

$H_0: p_1 - p_2 = 0$ $\quad$ $H_a: p_1 - p_2 < 0$

$H_0: p_1 - p_2 = 0$ $\quad$ $H_a: p_1 - p_2 \neq 0$

test stat

$$z^* = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

where

$$\hat{p} = \frac{X_1 + X_2}{n_1 + n_2}$$

successes in sample 1 $\quad$ # successes in sample 2

$$\left(\hat{p}_1 = \frac{X_1}{n_1}, \quad \hat{p}_2 = \frac{X_2}{n_2} \right)$$

$$\left[\frac{X_1 + X_2}{n_1 + n_2} \neq \frac{X_1}{n_1} + \frac{X_2}{n_2} \right]$$

Reference distribution

Standard normal

Requirement

$(n_1 \hat{p}_1 \geq 10$ and $n_1(1 - \hat{p}_1) \geq 10)$

AND

$(n_2 \hat{p}_2 \geq 10$ and $n_2(1 - \hat{p}_2) \geq 10)$

Large-sample confidence interval for comparing two proportions

Draw an SRS of size n_1 from a population having proportion p_1 of successes and draw an independent SRS of size n_2 from another population having proportion p_2 of successes. When n_1 and n_2 are large, an approximate level C confidence interval for $p_1 - p_2$ is

$$(\hat{p}_1 - \hat{p}_2) \pm z^* SE$$

In this formula the standard error SE of $\hat{p}_1 - \hat{p}_2$ is

$$SE = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

and z^* is the critical value for the standard Normal density curve with area C between $-z^*$ and z^*.

Assumptions and Conditions When Comparing Proportions

Independence Assumptions

Independent Response Assumption:

Within each group, we need independent responses from the cases. We cannot check that for certain, but randomization provides evidence of independence. So, we need to check the following:

- Randomization Condition: The data in each group should be drawn independently and at random from a population or generated by a completely randomized designed experiment.
- The 10% Condition: If the data are sampled without replacement, the sample should not exceed 10% of the population. If samples are bigger than 10% of the target population, random draws are no longer approximately independent.
- Independent Groups Assumption: The two groups we are comparing must be independent from each other.

Assumptions and Conditions When Comparing Proportions

Sample Size Condition
Each of the groups must be big enough. As with individual proportions, we need larger groups to estimate proportions that are near 0% and 100%. We check the success/failure condition for each group.

- Success/ Failure Condition: Both groups are big enough that at least 10 successes and at least 10 failures have been observed in each group or will be expected in each (when testing hypothesis).

Note: Two-sided significance tests (later we will discuss this concept) are robust against violations of this condition. In this case, we can conduct significance tests with smaller sample sizes. In practice, the two-sided significance test works well if there are at least five successes and five failures in each sample.

Hypotheses Tests for Two Proportions

To test the hypothesis $H_0 : p_1 = p_2$ first find the pooled proportion $\hat{p}$ of successes in both samples combined. Then compute the z_* statistic,
$z_* = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$
In terms of a variable Z having the standard Normal distribution, the approximate P-value for a test of H_0 against
$H_a : p_1 > p_2$: is : $P(Z > z_*)$
$H_a : p_1 < p_2$: is : $P(Z < z_*)$
$H_a : p_1 \neq p_2$: is : $2P(Z > |z_*|)$

Example

A hospital administrator suspects that the delinquency rate in the payment of hospital bills has increased over the past year. Hospital records show that the bills of 48 of 1284 persons admitted in the month of April have been delinquent for more than 90 days. This number compares with 34 of 1002 persons admitted during the same month one year ago. Do these data provide sufficient evidence to indicate an increase in the rate of delinquency in payments exceeding 90 days? Test using $\alpha = 0.10$.

Example (slide 112-113)

	① Current Yr	② Last Yr
	$n_1 = 1284$	$n_2 = 1002$
$\alpha = 0.10$	$x_1 = 48$	$x_2 = 34$
	$\hat{p}_1 = \frac{48}{1284} = \frac{4}{107}$	$\hat{p}_2 = \frac{34}{1002} = \frac{17}{501}$

$H_0: p_1 = p_2$

$H_a: p_1 - p_2 = 0$

$H_a: p_1 > p_2$

$H_a: p_1 - p_2 > 0$

Test stat

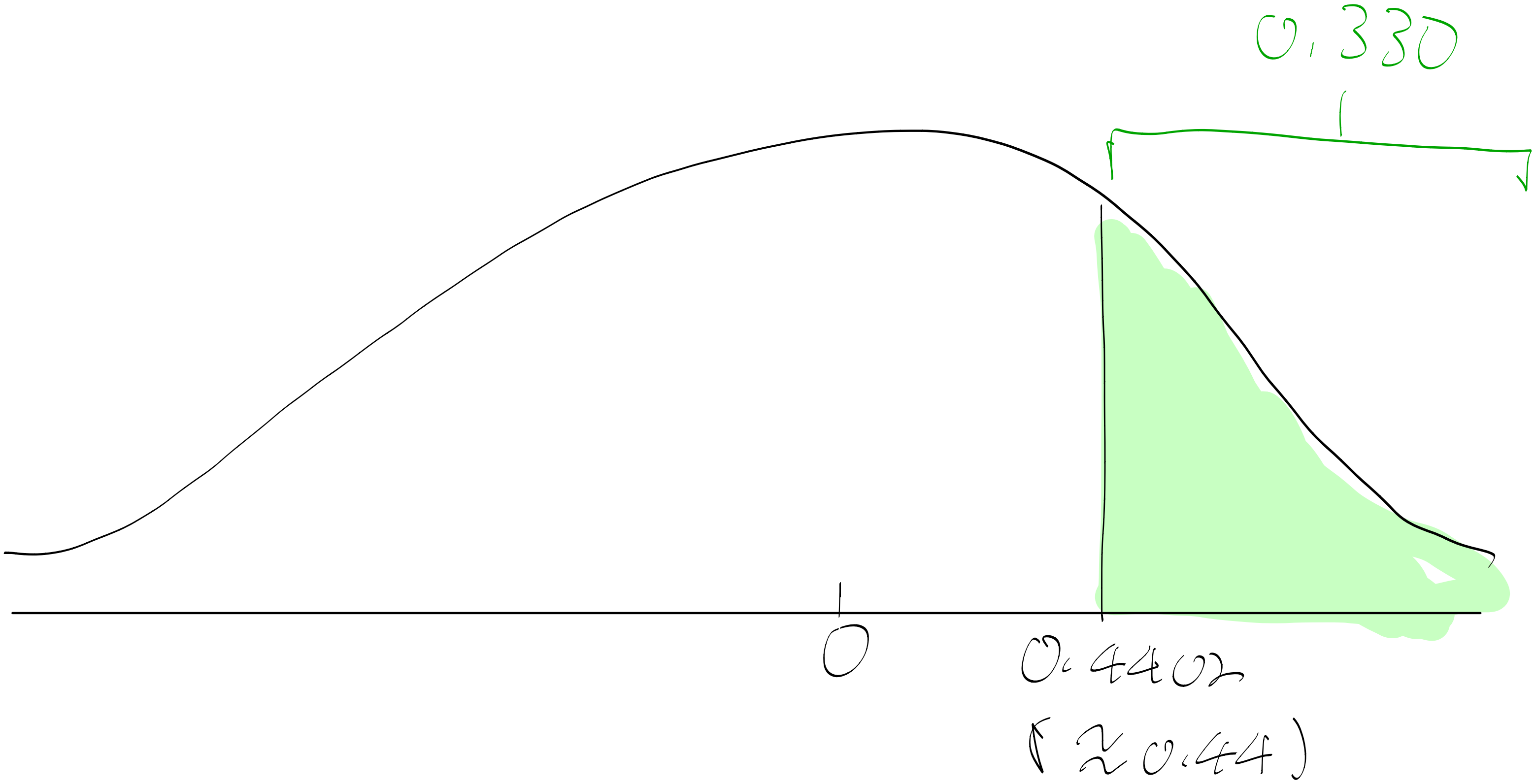
$$z^* = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$\hat{p} = \frac{X_1 + X_2}{n_1 + n_2}$$

$$\hat{p} = \frac{X_1 + X_2}{n_1 + n_2} = \frac{48 + 34}{1284 + 1002} = \frac{41}{1143}$$

$$z^* = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$= \frac{\left(\frac{4}{107} - \frac{17}{501}\right) - 0}{\sqrt{\frac{41}{1143} \cdot \left(1 - \frac{41}{1143}\right)\left(\frac{1}{1284} + \frac{1}{1002}\right)}} = 0.4402$$



p-val large. Insuff evidence to reject H0 that proportion of delinquencies is the same for both years.

Solution

Let p_1 and p_2 represent the proportions of all potential hospital admissions in April of this year and last year, respectively, that would have allowed their accounts to be delinquent for a period exceeding 90 days, and let $n_1 = 1284$ admissions this year and the $n_2 = 1002$ admissions last year represent independent random samples from these populations.

Solution

Step 1. State Hypotheses. $H_0 : p_1 = p_2$ vs $H_a : p_1 > p_2$
Step 2. Find test statistic. $\hat{p}_1 = \frac{48}{1284} = 0.0374$ and $\hat{p}_2 = \frac{34}{1002} = 0.0339$

$\hat{p} = \frac{x_1+x_2}{n_1+n_2} = \frac{48+34}{1284+1002} = 0.0359$

$z_* = \frac{\hat{p}_1-\hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = 0.45$

Solution

Step 3. Compute P-value.

$P - value = P(Z > z^*) = P(Z > 0.45) = 1 - P(Z < 0.45) = 0.3264$

Step 4. Conclusion. Since $P - value > \alpha = 0.10$, we **cannot** reject the null hypothesis that $p_1 = p_2$. The data present insufficient evidence to indicate that the proportion of delinquent accounts in April of this year exceeds the corresponding proportion last year.

Example: How to quit smoking

Nicotine patches are often used to help smokers quit. Does giving medicine to fight depression help? A randomized double-blind experiment assigned 244 smokers who wanted to stop to receive nicotine patches and another 245 to receive both a patch and the antidepression drug bupropion. Results: After a year, 40 subjects in the nicotine patch group had abstained from smoking, as had 87 in the patch-plus-drug group. How significant is the evidence that the medicine increases the success rate? State hypotheses, calculate a test statistic, use Table 6 to give its P-value, and state your conclusion. (Use $\alpha = 0.01$)

Solution

Step 1. State Hypotheses. $H_0 : p_1 = p_2$ vs $H_a : p_1 < p_2$
Step 2. Find test statistic. $\hat{p}_1 = \frac{40}{244} = 0.1639$ and $\hat{p}_2 = \frac{87}{245} = 0.3551$

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{40+87}{244+245} = 0.2597$$

$$z_* = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = -4.82$$

Solution

Step 3. Compute P-value.
$P - value = P(Z < z^*) = P(Z < -4.82) < 0.0003$

Step 4. Conclusion. Since $P - value < 0.0003 < \alpha = 0.01$, we **reject** the null hypothesis that $p_1 = p_2$. The data provide very strong evidence that bupropion increases success rate.

R Code

```
successes=c(87, 40);

totals=c(245, 244);

prop.test(successes,totals, alternative="greater",
correct=FALSE);
```

R Code

$z = 4.62$

```
##
##  2-sample test for equality of proportions without
##  continuity correction
##
## data:  x and n
## X-squared = 23.237, df = 1, p-value = 7.161e-07
## alternative hypothesis: greater
## 95 percent confidence interval:
##  0.1275385 1.0000000
## sample estimates:
##    prop 1    prop 2
## 0.3551020 0.1639344
```

How does χ^2 in R output relate to Z stat?

$$\chi^{*2} = (Z^*)^2 \qquad Z^* = \pm\sqrt{\chi^2}$$

Why

$$Z^2 \sim \chi^2_{(1)}$$

Comparing Two Proportions: Epidural and Nursing At Six Months

There is some concern that if a woman has an epidural to reduce pain during childbirth, the drug can get into the baby's bloodstream, making the baby sleepier and less willing to breastfeed. In 2006, the International Breastfeeding Journal published results of a study conducted at Sydney University. Researchers followed up on 1178 births, noting whether the mother had an epidural and whether the baby was still nursing after six months. The results are summarized in a contingency table.

Do breastfeeding proportions differ between mothers who had epidural and those who did not? Let p_1 denote the proportion among mothers that had epidural who are breastfeeding at 6 months. Let p_2 denote the proportion among mothers that did not have epidural who are breastfeeding at 6 months.

Data

	Breasfeeding at 6 Months	
Epidural	Yes	No
Yes	206	190
No	498	284

Checking Assumptions and Conditions

Randomization Condition:
We do not know whether mother were randomly selected, but we can view them as representative of a larger collection of mothers under similar conditions.
Independent Groups Assumption:
It is reasonable to believe that mothers who had epidural and mother who did not have epidural to reduce pain during birth are independent of each other. **10% Condition:** We can imagine many more mothers under similar conditions.
Success/Failure Condition: For mothers who had epidural, the count for successes was 206, and for failure was 190; For mothers who did not have epidural the count for successes was 498, for failure was 284; The observed numbers of both success and failures are more than 10 for both groups.
Since these conditions are met, we can use a two-proportion Z-CI.

Data (again)

Conditional Percentages are also displayed in each cell.

	Breasfeeding at 6 Months	
Epidural	Yes	No
Yes	206	190
	52%	48%
No	498	284
	64%	36%

Is there a difference in proportion of breastfeeding mothers who had epidural and those who did not?
Let p_1 denote the proportion among mothers that had epidural who are breastfeeding at 6 months. Let p_2 denote the proportion among mothers that did not have epidural who are breastfeeding at 6 months.
$\hat{p}_1 \approx \frac{206}{396} = 0.52$ and $\hat{p}_2 \approx \frac{498}{782} = 0.64$.

Solution

Step 1. State Hypotheses. $H_0 : p_1 = p_2$ vs $H_a : p_1 \neq p_2$
Step 2. Find test statistic. $\hat{p}_1 = \frac{206}{396} = 0.52$ and $\hat{p}_2 = \frac{498}{782} = 0.64$

$\hat{p} = \frac{x_1+x_2}{n_1+n_2} = \frac{206+498}{396+782} = 0.5976$

$z_* = \frac{\hat{p}_1-\hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = -3.86$

Solution

Step 3. Compute P-value.
$P - value = 2P(Z > |z^*|) = 2P(Z > |-3.86|) \approx 0.0001$

Step 4. Conclusion. Since $P - value$ is very small, we **reject** the null hypothesis that $p_1 = p_2$. The data provide very strong evidence that mothers who had epidural are less likely to breastfeed at 6 months than those who did not have epidural.

R Code

```
successes=c(206, 498);

totals=c(396, 782);

prop.test(successes,totals, correct=FALSE);
```

R Code

```
##
##  2-sample test for equality of proportions without
##  continuity correction
##
## data:  x and n
## X-squared = 14.869, df = 1, p-value = 0.0001152
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  -0.17626992 -0.05698333
## sample estimates:
##    prop 1    prop 2
## 0.5202020 0.6368286
```

Follow-up

We are 95% confident the percent breastfeeding mothers at 6 months for those who had epidural are between 5.69% and 17.62% less than those who did not have epidural.

Comparing Variances

Comparing Two Population Variances: Independent Sampling

How do you know whether the homogeneity of variance assumption is satisfied?
One simple method involves just looking at two sample variances. Logically, if two population variances are equal, then the two sample variances should be very similar. When the two sample variances are reasonably close, you can be reasonably confident that the homogeneity assumption is satisfied and proceed with, for example, Student t-interval. However, when one sample variance is three or four times larger than the other, then there is reason for a concern. The common statistical procedure for comparing population variances σ_1^2 and σ_2^2 makes an inference about the ratio of $\frac{\sigma_1^2}{\sigma_2^2}$.

Making An Inference for Ratio of Population Variances

To make an inference about the ratio of $\frac{\sigma_1^2}{\sigma_2^2}$ we collect sample data and use the ratio of the sample variances $\frac{s_1^2}{s_2^2}$.

The sampling distribution of $\frac{s_1^2}{s_2^2}$ is based on the two of the assumptions already required for the t procedure:

1. The two sampled populations are Normally distributed.
2. The samples are Normally and independently selected from their respective populations.

When these assumptions are satisfied, the sampling distribution of $\frac{s_1^2}{s_2^2}$ is an F-distribution with $(n_1 - 1)$ numerator degrees of freedom and $(n_2 - 1)$ denominator degrees of freedom.

Hypothesis Tests for Two Variances

We know that $\frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2} = \frac{\sigma_2^2 s_1^2}{\sigma_1^2 s_2^2} \sim F_{n_1-1,n_2-1}$.

Null hypothesis: $H_0 : \sigma_1^2 = \sigma_2^2$

Test statistic: $F^* = \frac{s_1^2}{s_2^2}$ set up with $s_1^2 > s_2^2$ (for tables - it's easier!!)

Decision rules

$H_a: \sigma_1^2 \neq \sigma_2^2$ reject H_0 if $F^* > F_{n_1-1,n_2-1,\alpha/2}$ or $F^* < F_{n_1-1,n_2-1,1-\alpha/2}$.
$H_a: \sigma_1^2 > \sigma_2^2$ reject H_0 if $F^* > F_{n_1-1,n_2-1,\alpha}$ OR $P(F_{n_1-1,n_2-1} > F^*)$ is too small.
$H_a: \sigma_1^2 < \sigma_2^2$ reject H_0 if $F^* < F_{n_1-1,n_2-1,1-\alpha}$ OR $P(F_{n_1-1,n_2-1} < F^*)$ is too small.

Comparing Two Population Variances Managerial Success Indexes for Two Groups.

Behavioural researchers have developed an index designed to measure managerial success. The index (measured on a 100-point scale) is based on the manager's length of time in the organization and their level within the term; the higher the index, the more successful the manager. Suppose a researcher wants to compare the average index for the two groups of managers at a large manufacturing plant. Managers in group 1 engage in high volume of interactions with people outside the managers' work unit (such interaction include phone and face-to-face meetings with customers and suppliers, outside meetings, and public relation work). Managers in group 2 rarely interact with people outside their work unit.

Example

Independent random samples of 12 and 15 managers are selected from groups 1 and 2, respectively, and success index of each is recorded. Note: The response variable is "Managerial Success Indexes".

- Managerial success indexes is a continuous quantitative variable, measured on 100-point scale.

The explanatory variable is "Type of group".

- Type of group (Group 1: Interaction with outsiders, Group 2: Fewer interactions) is a nominal categorical variable.

R Code

```
# Importing data file into R;

success=read.csv(file="success.csv",header=TRUE);

# Getting names of variables;

names(success);

# Attaching data file;
attach(success);
```

R Code

```
## [1] "Success_Index" "Group"
```

R Code (Descriptive Statistics)

```
# loading library mosaic;

library(mosaic);

favstats(Success_Index~Group);
```

Note. Group 1 = "interaction with outsiders" and Group 2 = "fewer interaction".

R Code (Descriptive Statistics)

```
##    .group min    Q1 median    Q3 max     mean       sd  n mi
## 1       1  53 62.25   65.5 69.25  78 65.33333 6.610368 12
## 2       2  34 42.50   50.0 54.50  68 49.46667 9.334014 15
```

Note. Group 1 = "interaction with outsiders" and Group 2 = "fewer interactions".

$H_0 : \frac{\sigma_1^2}{\sigma_2^2} = 1$ versus $H_a : \frac{\sigma_1^2}{\sigma_2^2} \neq 1$
Test statistic: $F^* = \frac{s_1^2}{s_2^2} = \frac{6.6103^2}{9.3340^2} = 0.50155$
Reference distribution: $F_{n_1-1=11, n_2-1=14}$
Significance level: $\alpha = 0.05$.

R Code (Critical values)

```r
# Finding F-critical value with R
# alpha = 0.05;
# alpha/2 = 0.025;

qf(0.025, df1=11, df2=14);

## [1] 0.2977245

qf(0.975, df1=11,df2=14);

## [1] 3.09459
```

95% Confidence Interval for $\frac{\sigma_1^2}{\sigma_2^2}$

$$\left(\frac{6.610368^2}{9.334014^2(3.0945898)}, \frac{6.610368^2}{9.334014^2(0.2977245)}\right)$$

$$(0.1621, 1.6846)$$

Since "1" is in this 95% CI, we have no evidence that the population variances of managerial success indexes for the two groups differ.

R Code (Critical values)

```
# 95% CI for the ratio of two variances;

var.test(Success_Index~Group);

##
##  F test to compare two variances
##
## data:  Success_Index by Group
## F = 0.50155, num df = 11, denom df = 14, p-value = 0.2554
## alternative hypothesis: true ratio of variances is not equa
## 95 percent confidence interval:
##  0.1620733 1.6846122
## sample estimates:
## ratio of variances
##          0.5015503
```