15.075 Spring 2003
Review for Mid-term

Slides prepared by Elizabeth Newton.

As a statistician…

- You work for a company developing a new non-invasive method for measuring cardiac output.
- Research Questions
- Study Design
- Sampling Design
- Sample Size

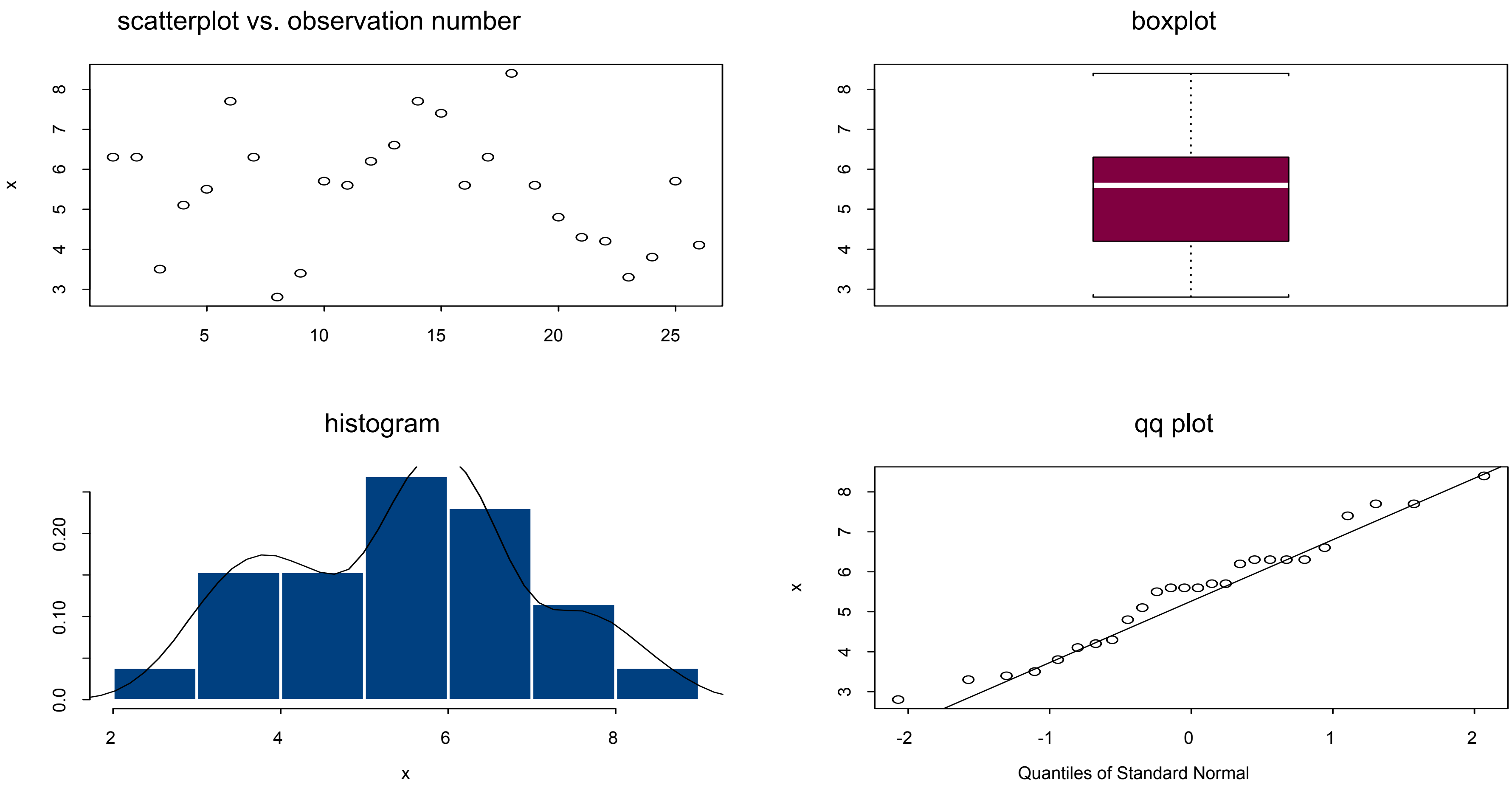
Sample Size

- $n = [(z_{\alpha/2}+z_{\beta})\sigma_D/\delta]^2$
- $\alpha=0.01$
- $\beta=0.1$
- $\sigma/\delta = 2$
- $n = [(2.58+1.28)2]^2 = 60$

Cardiac Output Data

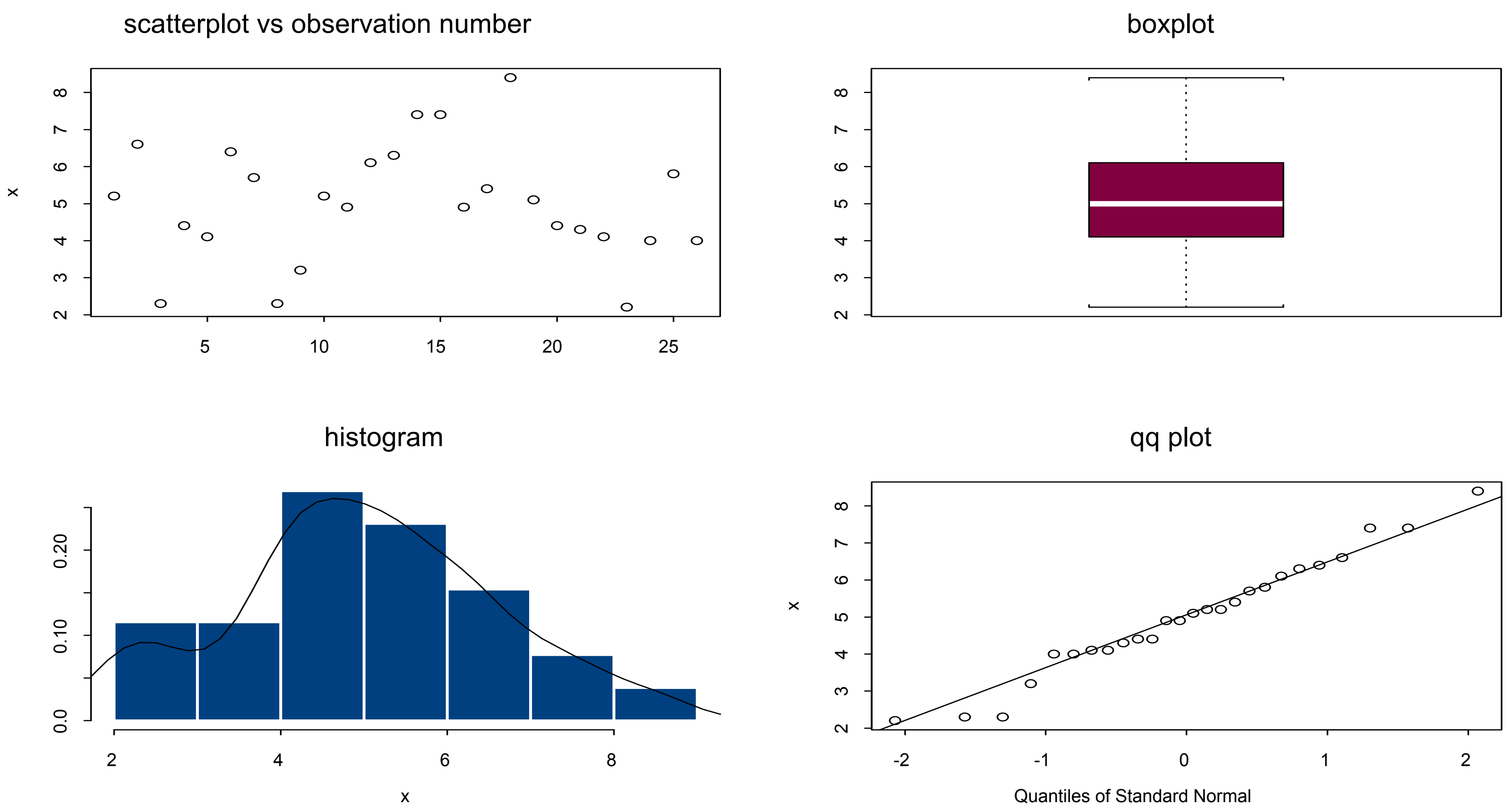
(See Table 8.6 on page 285 of the course textbook.)

Cardiac Output - Method A



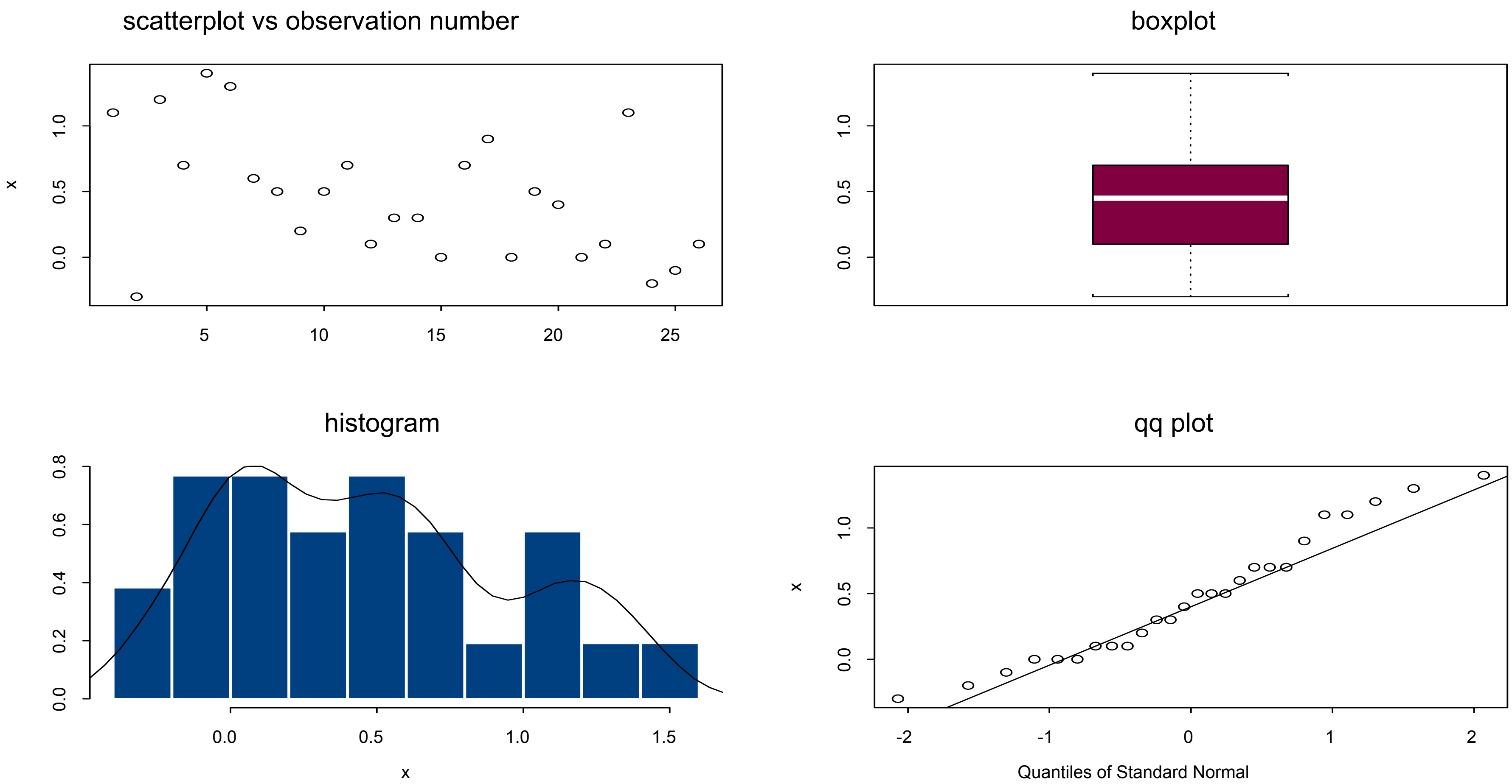
These graphs were created using S-PLUS(R) Software. S-PLUS(R) is a registered trademark of Insightful Corporation.

Cardiac Output - Method B



These graphs were created using S-PLUS(R) Software. S-PLUS(R) is a registered trademark of Insightful Corporation.

Cardiac Output - Differences



These graphs were created using S-PLUS(R) Software. S-PLUS(R) is a registered trademark of Insightful Corporation.

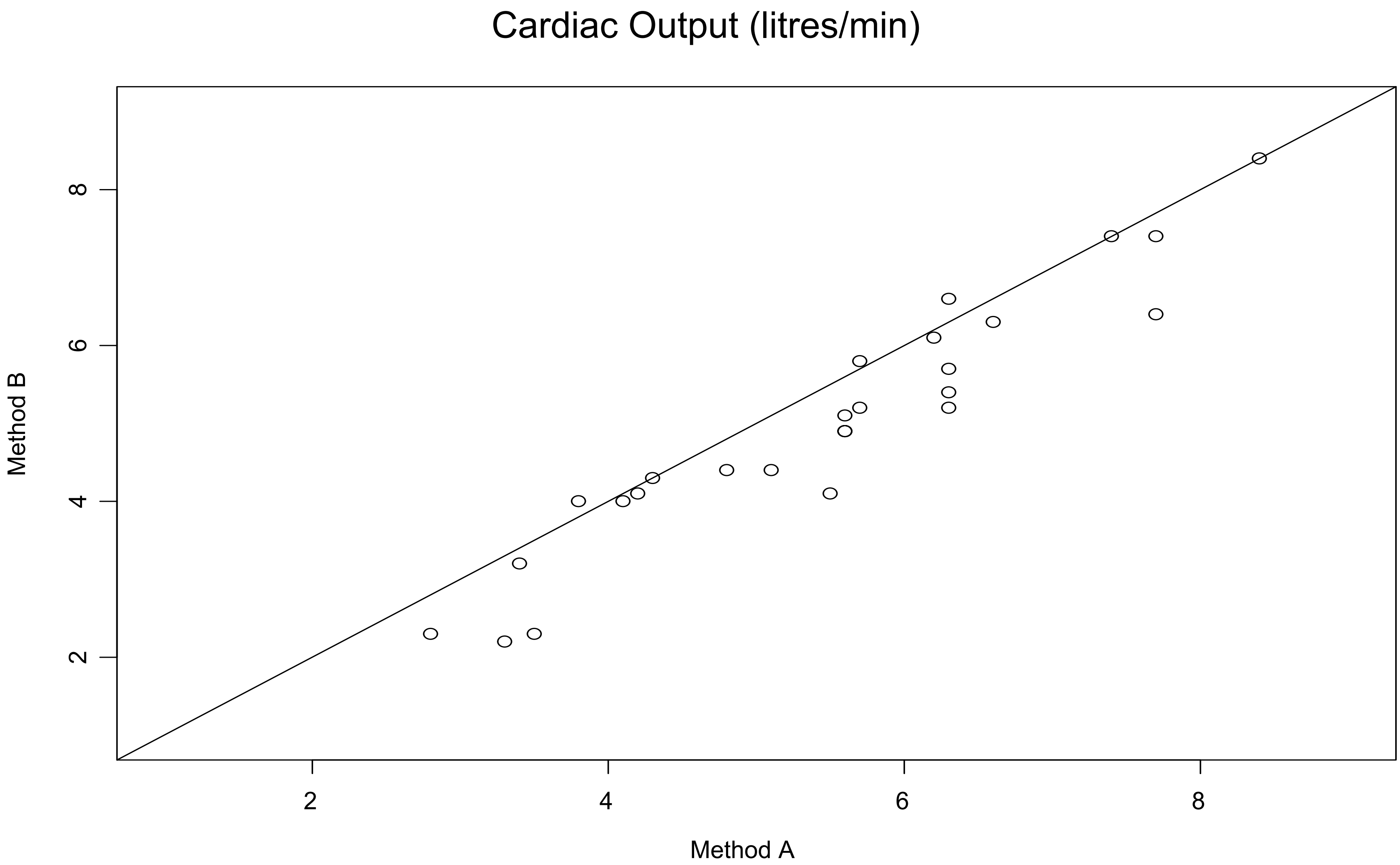
Data Summaries

```
fsummary(A)

$quantile:
  0%    25%   50%   75%  100%
 2.8  4.225   5.6    6.3   8.4

$mean:
[1] 5.469231

$stdev:
[1] 1.486141

$iqr:
[1] 2.075

$range:
[1] 5.6

$n
[1] 26

$nmiss:
[1] 0
```

```
fsummary(B)

$quantile:
  0%   25%   50%    75%  100%
 2.2   4.1    5    6.025  8.4

$mean:
[1] 5.003846

$stdev:
[1] 1.579742

$iqr:
[1] 1.925

$range:
[1] 6.2

$n:
[1] 26

$nmiss:
[1] 0
```

```
fsummary(Di)

$quantile:
  0%   25%   50%   75%  100%
-0.3   0.1   0.45   0.7   1.4

$mean:
[1] 0.4653846

$stdev:
[1] 0.4824457

$iqr:
[1] 0.6

$range:
[1] 1.7

$n:
[1] 26

$nmiss:
[1] 0
```

Example 8.6, page 285



Confidence Interval

95% confidence interval on difference:

û ± cd, where

û=estimated difference,

c=critical constant ($t_{\alpha/2,n-1}$),

d=standard deviation ($s/\sqrt{n}$)

0.465 ± 2.06 * 0.482 / 5.1 = {0.271, 0.660}

Conclusion?

t-statistic and P-value

$T=(\hat{u}-\delta_0)/s$

$=(0.465-0) / (0.482/ 5.1)$

$= 4.92$

$> 2.06 = t_{\alpha/2,25}$

P-value = 2*(1-pt(4.92, 25))=0.000046

What if measurements were on different patients?

Assume variances equal:

Since sample sizes are the same (eqn 8.6, page 275):

T=(mean(A)-mean(B))/sqrt((var(A)+var(B))/26)

=1.0941 < 2.01 = qt(0.975,50)

P-value = 2*(1-pt(1.0941,50)) = 0.279

Assume variances are not equal:

Calculate df = $25*(w_1+w_2)^2/(w_1^2+w_2^2)$, where w1 and w2 are the standard errors of the means of A and B (eqn. 8.11, page 280)

df=49.81

Results are almost the same as above.

Why so different?

Independent:

```
Tstat<-(mean(A)-mean(B))/sqrt((var(A)+var(B))/26)
```

(Tstat=1.094097)

Matched:

```
Tstat<-(mean(A)-mean(B))/sqrt((var(A)+var(B)-2*var(A,B))/26)
Tstat<-(mean(A)-mean(B))/sqrt(var(A-B)/26)
```

(Tstat= 4.918699)

A different problem

- Suppose, instead, that these 2 methods are designed to boost cardiac output.

- Interest is in comparing proportion of patients with cardiac output greater than 5.4 liters/minute.

Observed

Method	>5.4	≤5.4	Total
A	16	10	26
B	9	17	26
Total	25	27	52

Expected

Method	>5.4	≤5.4	Total
A	12.5	13.5	26
B	12.5	13.5	26
Total	25	27	52

Independent Samples Design

Test using Chi-square or z statistic (with pooled variance).

$\chi^2 = \Sigma_i (O_i - E_i)^2/E_i$

$= (16-12.5)^2/12.5 + (9-12.5)^2/12.5 + (10-13.5)^2/13.5 + (17-13.5)^2/13.5$

$=3.7748$

p1= 16/26 = 0.615

p2 = 9/26 = 0.366

p = 25/52 = 0.481

z<-(p1-p2)/sqrt(p*q*(2/52)) = 1.943

$z^2 = \chi^2$

P-value = 0.052