Slide 1

Back Of Envelope Statistical Calculations

Rudy Kittlitz
2007 March 14

Slide 2

Quick And Dirty Statistics

- Need to make quick calculations
- Must be portable
- Must have confidence in conclusions
- Will review
  - Outside Count test
  - Corner Count test
  - Line of Best Fit
- Must be able to count and to remember 7 & 11

Slide 3

Outside Count Test For The Difference In Averages

- New fiber 'B' proposed for market now dominated by fiber 'A'
  - To justify move to next step, fiber 'B' must be stronger than fiber 'A' by at least one unit.
  - Six laboratory batches of each fiber's results
    - A: 4.9, 3.0, 6.5, 7.3, 4.6, 4.1
    - B: 5.0, 9.0, 6.9, 4.2, 7.2, 6.5
- The optimist presents his case

Slide 4

Outside Count Test For The Difference In Averages

DATA

	Fiber 'A'	Fiber 'B'
	3.0	4.2
	4.1	5.0
	4.6	6.5
	4.9	6.9
	6.5	7.2
	7.3	9.0
Averages	5.1	6.5
Difference	1.4	
Req'd Diff.	1.0	

Slide 5

Outside Count Test For The Difference In Averages

- The pessimist present his analysis of the data by sorting the data first

Fiber 'A'	Fiber 'B'
3.0	
4.1	
	4.2
4.6	
4.9	
	5.0
6.5	6.5
	6.9
	7.2
7.3	
	9.0

- Now count the number of data in 'separate tails'

Slide 6

Outside Count Test For The Difference In Averages

- Add the two counts: 2 + 1 = 3
- If count > 7 Samples are different with 95% confidence [2-sided test]
- If count < 7 Samples not different
- Sample size $4 \le n_1 \le n_2 \le 30$
  $n_2 \le 3 + 4n_1/3$
- [John Tukey 1959]

Slide 7

Is Fiber 'B' Different Than Fiber 'A'?

Two-Sample T-Test and CI: B, A

```
   N  Mean  StDev  SE Mean
B  6  6.47   1.70     0.69
A  6  5.07   1.58     0.65

Difference = mu (B) - mu (A)
Estimate for difference:  1.40
95% CI for difference:  (-0.74315, 3.54315)

T-Test of difference = 0 (vs not =): T-Value
  = 1.48  P-Value = 0.174  DF = 9
```

Slide 8

Corner Count Test

- Is there a correlation?
- Corner Count test can answer question if:
  - You can count
  - You can use a straight-edge
  - [John Tukey 1947]

Slide 9

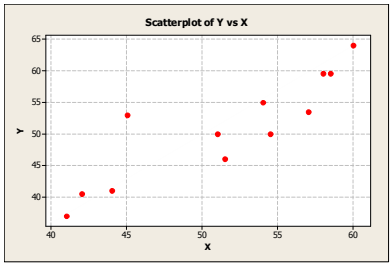
Corner Count Test
Data

Row	X	Y
1	41.0	37.0
2	42.0	40.5
3	44.0	41.0
4	45.0	53.0
5	51.0	50.0
6	51.5	46.0
7	54.0	55.0
8	54.5	50.0
9	57.0	53.5
10	58.0	59.5
11	58.5	59.5
12	60.0	64.0

Slide 10

Corner Count Test

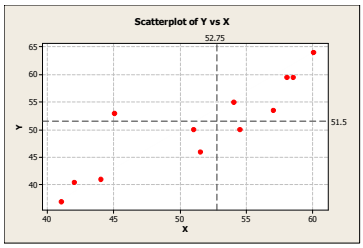
- Plot the data



Slide 11

Corner Count Test

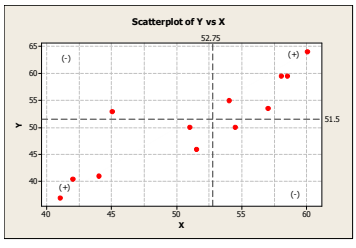
- Determine the X and Y medians



Slide 12

Corner Count Test

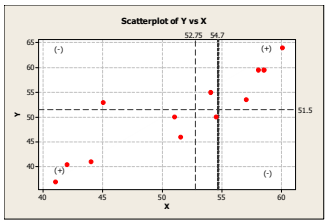
- Label quadrant with '+' and '-' as in analytic geometry



Slide 13

Corner Count Test

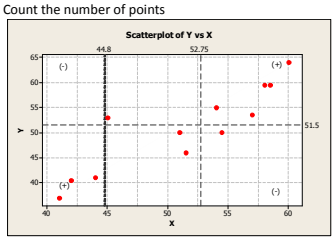
- Move a vertical line from Right border toward the Left until a point in the other quadrant is encountered
- Count the number of points



Slide 14

Corner Count Test

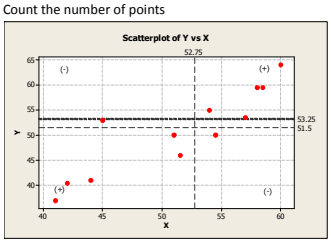
- Move a vertical line from Left border toward the Right until a point in the other quadrant is encountered
- Count the number of points



Slide 15

Corner Count Test

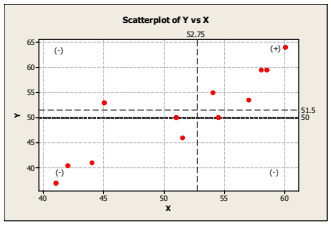
- Move a horizontal line from Top border toward the Bottom until a point in the other quadrant is encountered
- Count the number of points



Slide 16

Corner Count Test

- Move a horizontal line from Bottom border toward the Top until a point in the other quadrant is encountered
- Count the number of points



Slide 17

Corner Count Test

- Count the number of points as follows
  - R > L +4
  - L > R +3
  - T > B +5
  - B > T +4 ½
  - Total 16 ½
- If absolute sum >/= 11, then 95% confident that there is a correlation

Slide 18

Corner Count Test

- Correlation coefficient, r, = 0.886 [significant at p = 0.000]
- Quick estimate to check if "r" is significant at p = 0.05

n	df	2/SR[n]	r@0.05
3	1	1.155	0.997
4	2	1.000	0.950
5	3	0.894	0.878
6	4	0.816	0.811
7	5	0.756	0.754
8	6	0.707	0.707
9	7	0.667	0.666
10	8	0.632	0.632
11	9	0.603	0.602
12	10	0.577	0.576
13	11	0.555	0.553
14	12	0.535	0.532
15	13	0.516	0.514

Slide 19

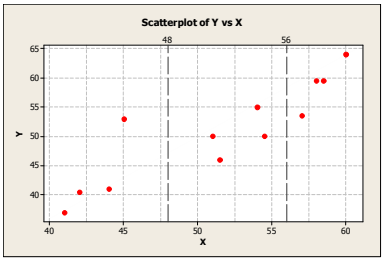
Line of Best Fit

- If there is a correlation, can then determine "Line of Best Fit" that is equivalent to Least-Squares equation
  - Must be able to count
  - Able to use a straight-edge
  - [M. H. Quenouille 1959]

Slide 20

Line of Best Fit

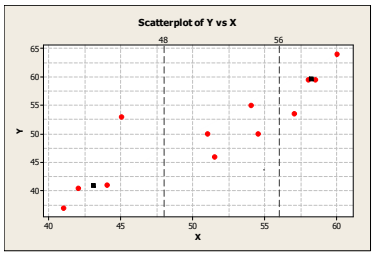
- Plot the data and divide data with vertical lines into 3 equal groups



Slide 21

Line of Best Fit

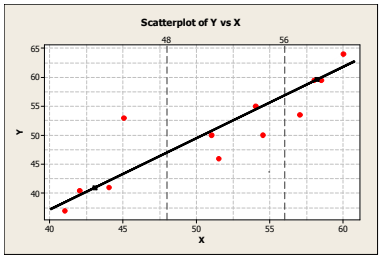
- Determine X and Y medians for 'end' groups



Slide 22

Line of Best Fit

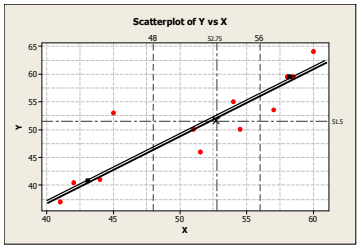
- Connect intersection of 'end' group medians



Slide 23

Line of Best Fit

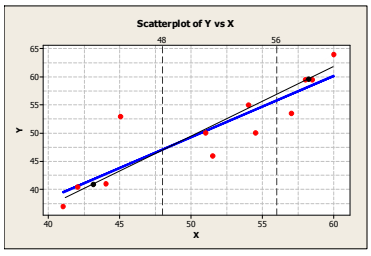
- Improve estimate by dropping perpendicular line to original medians and moving initial line 1/3rd toward original medians



Slide 24

Line of Best Fit

- "Line of Best Fit" compared to "Least-squares Line [blue line]



Slide 25

Conclusions

- Three portable methods illustrated that are easy to use
- They are distribution-free [don't need to know the underlying distribution]
- Have to be able to count
- Have to remember 7 & 11

Slide 26

Estimating Standard Deviation

41	45	47	52	53	47	52	47	46
48	47	43	51	52	56	44	47	46
53	46	44	49	53	56	48	49	48
56	50	46	49	48	46	54	54	55

For the data in row 1 compute the value of range/sqrt(n) = 4
[actual standard deviation = 3.898005]

For the data in row 2, 3, and 4 compute the range for each row: 13, 12, 10

Then for the last 3 rows of data, compute the estimated standard deviation = Σ (r) / 3*sqrt(n) = 35/3*sqrt(9) = 3.88
[actual standard deviation = 3.945137]