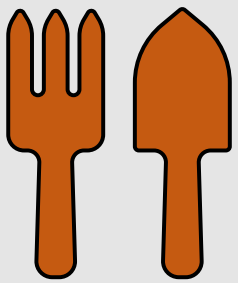
Getting Your Hands Dirty with Statistics

BERDC Special Topics Talk 9



DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

Dr. Mark Williamson

Biostatistics, Epidemiology, and Research Design Core

DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

Opening

UND
UNIVERSITY OF NORTH DAKOTA.

Goal: Run standard statistical tests by hand

- Compacted history of statistics
- Chart of tests that can be done by hand
- Collecting our own data
- Chi-Square
- T-test
- ANOVA
- Correlation
- Regression

Before Moving On:

Pre-test: https://und.qualtrics.com/jfe/form/SV_8Cgkbcmu5HSVQVo

DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

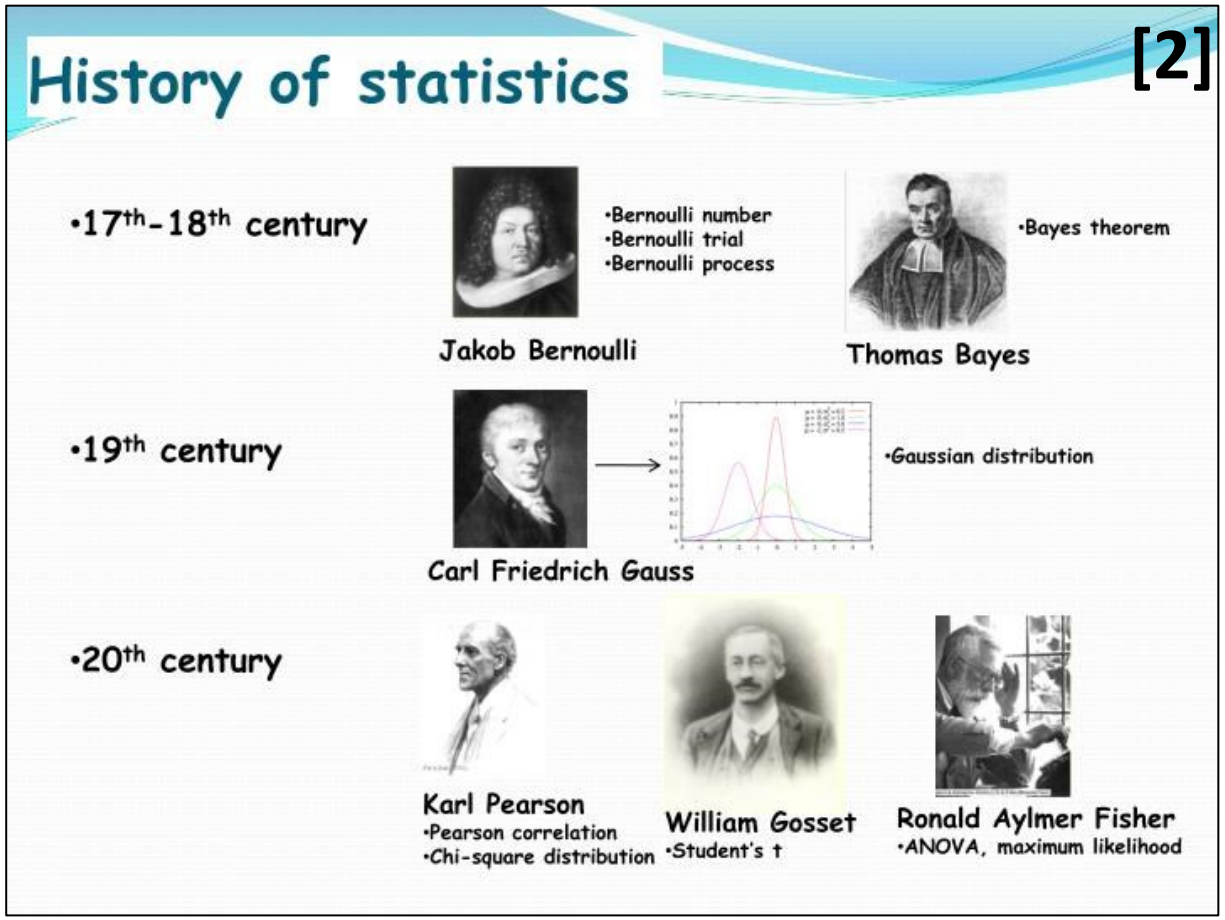
History

UND UNIVERSITY OF NORTH DAKOTA.

History of Statistics [1]

- 17th century
  - Probability theory
- 18th century
  - Inferential statistics
- 19th century
  - Statistics applied to education and sociology
- 20th century
  - Regression and correlation; computer analysis

History of statistics [2]



DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

History 2

UND UNIVERSITY OF NORTH DAKOTA.

[3]

History of Statistics Timeline

20th Century (early)	Pearson	studied natural selection using correlation, formed first academic department of statistics, Biometrika journal, helped develop the Chi Square analysis
	Gossett (Student)	studied process of brewing, alerted the statistics community about problems with small sample sizes, developed Student's test
	Fisher	evolutionary biologists - developed ANOVA, stressed the importance of experimental design
20th Century (later)	Wilcoxon	biochemist studied pesticides, non parametric equivalent of two-samples test
	Kruskal, Wallis	economists who developed the non parametric equivalent of the ANOVA
	Spearman	psychologist who developed a non-parametric equivalent of the correlation coefficient
	Kendall	statistician who developed another non parametric equivalent the correlation coefficient
	Tukey	statistician who developed multiple comparisons procedure
	Dunnett	biochemist who studied pesticides, developed multiple comparisons procedure for control groups
	Keuls	agronomist who developed multiple comparisons procedure
	Computer Technology	provided many advantages over calculations by hand or by calculator, stimulated the growth of investigation into new techniques

DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

History 2

UND UNIVERSITY OF NORTH DAKOTA.

[3]

History of Statistics Timeline

20th Century (early)	Pearson	studied natural selection using correlation, formed first academic department of statistics, Biometrika journal, helped develop the Chi Square analysis
	Gossett (Student)	studied process of brewing, alerted the statistics community about problems with small sample sizes, developed Student's test
	Fisher	evolutionary biologists - developed ANOVA, stressed the importance of experimental design

20th Century (later)	Wilcoxon	biochemist studied pesticides, non parametric equivalent of two-samples test
	Kruskal, Wallis	economists who developed the non parametric equivalent of the ANOVA
	Spearman	psychologist who developed a non-parametric equivalent of the correlation coefficient
	Kendall	statistician who developed another non parametric equivalent the correlation coefficient
	Tukey	statistician who developed multiple comparisons procedure
	Dunnett	biochemist who studied pesticides, developed multiple comparisons procedure for control groups
	Keuls	agronomist who developed multiple comparisons procedure
	Computer Technology	provided many advantages over calculations by hand or by calculator, stimulated the growth of investigation into new techniques

We won't be using (much) computer technology today

DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

Cans and Cants

UND
UNIVERSITY OF NORTH DAKOTA

Test	Feasibility
T-tests	
1-sample T-test	
2-sample T-test	
Paired T-test	
Frequencies	
Fisher's exact test	
Chi-Squared	
ANOVA	
1-way ANOVA	
2-way ANOVA	
≥3-way ANOVA	
Repeated measures ANOVA	
Nested/Block ANOVA	

Test	Feasibility
Regression and Correlation	
Pearson Correlation	
Simple linear regression	
Multiple linear regression	
Generalized linear models	
Linear mixed models	
Generalized linear mixed models	
Assorted	
Multivariate Analysis	
Bayesian Analysis	
Survival Analysis	
Longitudinal Analysis	
Etc.	

DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

Cans and Cants

UND
UNIVERSITY OF NORTH DAKOTA

Test	Feasibility
T-tests	
1-sample T-test	**Yes**
2-sample T-test	**Yes**
Paired T-test	**Yes**
Frequencies	
Fisher's exact test	**Yes**
Chi-Squared	**Yes**
ANOVA	
1-way ANOVA	**Yes**
2-way ANOVA	**Yes**
≥3-way ANOVA	**Yes**
Repeated measures ANOVA	**Yes, but hard**
Nested/Block ANOVA	**Yes, but hard**

Test	Feasibility
Regression and Correlation	
Pearson Correlation	**Yes**
Simple linear regression	**Yes**
Multiple linear regression	**Yes, but hard**
Generalized linear models	**No**
Linear mixed models	**No**
Generalized linear mixed models	**No**
Assorted	
Multivariate Analysis	**No**
Bayesian Analysis	**No**
Survival Analysis	**No**
Longitudinal Analysis	**No**
Etc.	**No**

DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

Cans and Cants

UND UNIVERSITY OF NORTH DAKOTA

Test	Feasibility
T-tests	
1-sample T-test	**Yes**
2-sample T-test	**Yes**
Paired T-test	**Yes**
Frequencies	
Fisher's exact test	**Yes**
Chi-Squared	**Yes**
ANOVA	
1-way ANOVA	**Yes**
2-way ANOVA	**Yes**
≥3-way ANOVA	**Yes**
Repeated measures ANOVA	**Yes, but hard**
Nested/Block ANOVA	**Yes, but hard**

Test	Feasibility
Regression and Correlation	
Pearson Correlation	**Yes**
Simple linear regression	**Yes**
Multiple linear regression	**Yes, but hard**
Generalized linear models	**No**
Linear mixed models	**No**
Generalized linear mixed models	**No**
Assorted	
Multivariate Analysis	**No**
Bayesian Analysis	**No**
Survival Analysis	**No**
Longitudinal Analysis	**No**
Etc.	**No**

We'll cover five standard tests today (in green)

DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

Homegrown Data

UND
UNIVERSITY OF NORTH DAKOTA.

Chi-Square
- Color (L/D) across 2 locations*

T-test
- Length across 2 locations

ANOVA
- Length across 3 locations

Correlation/Regression
- Length versus width

Wood Chips

**more samples should normally be used (small samples are problematic for Chi Squared and Exact is better)*

DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

Homegrown Data 2

UND
UNIVERSITY OF NORTH DAKOTA.

Section A

Section B

Section C

DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

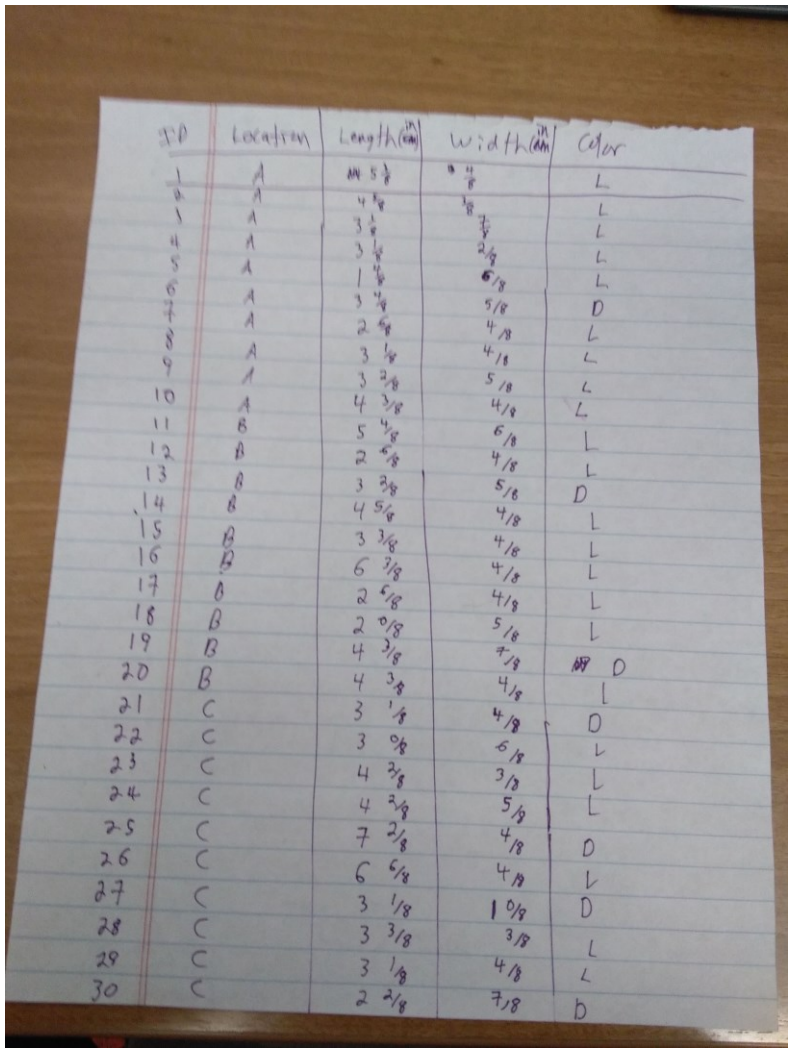
Homegrown Data 3

UND
UNIVERSITY OF NORTH DAKOTA.

DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

Homegrown Data 4

UND
UNIVERSITY OF NORTH DAKOTA.



DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

Homegrown Data 4

UND
UNIVERSITY OF NORTH DAKOTA

ID	Location	Length	Width	Color
1	A	5.375	0.5	L
2	A	4.625	0.375	L
3	A	3.125	0.875	L
4	A	3.125	0.25	L
5	A	1.5	0.75	L
6	A	3.5	0.625	D
7	A	2.75	0.5	L
8	A	3.125	0.5	L
9	A	3.25	0.625	L
10	A	4.375	0.5	L

ID	Location	Length	Width	Color
11	B	5.5	0.75	L
12	B	2.75	0.5	L
13	B	3.25	0.625	D
14	B	4.625	0.5	L
15	B	3.375	0.5	L
16	B	6.375	0.5	L
17	B	2.75	0.5	L
18	B	2	0.625	L
19	B	4.375	0.875	D
20	B	4.375	0.5	L

ID	Location	Length	Width	Color
21	C	3.125	0.5	D
22	C	3	0.75	L
23	C	4.25	0.375	L
24	C	4.25	0.625	D
25	C	7.25	0.5	L
26	C	6.75	0.5	L
27	C	3.125	1	D
28	C	3.375	0.375	L
29	C	3.125	0.5	L
30	C	2.25	0.875	D

DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

Chi-Squared

[4]

UND
UNIVERSITY OF NORTH DAKOTA.

Steps

1. State hypothesis
2. Select alpha
3. Fill out table
4. Calculate observed values (O)
5. Calculate expected values (E)
6. Calculate subgroup values
7. Calculate test statistic (χ^2)
8. Calculate degrees of freedom (df)
9. Look up test statistic in table
10. State conclusion

$$\chi^2 = \sum_{i,j} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

df=(# rows−1) x (# columns−1)

DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

Chi-Squared 2

UND UNIVERSITY OF NORTH DAKOTA.

	Total	A	B	C
L	23	9	8	6
D	7	1	2	4

Steps

1. **State hypothesis**
2. **Select alpha**
3. **Fill out table**
4. **Calculate observed values (O)**
5. **Calculate expected values (E)**
6. **Calculate subgroup values**
7. Calculate test statistic (χ^2)
8. Calculate degrees of freedom (df)
9. Look up test statistic in table
10. State conclusion

H_0

NO significant relationship b/t location and color

H_1

significant relationship

$\alpha = 0.05$

$$\chi^2 = \sum_{i,j} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

	A	B	
L	9	8	17
D	1	2	3
	10	10	20

A/B	O	E	Sub-group
L A	9	(17·10)/20 = 8.5	(9-8.5)²/8.5 = 0.029
D A	1	(3·10)/20 = 1.5	(1-1.5)²/1.5 = 0.167
L B	8	(17·10)/20 = 8.5	(9-8.5)²/8.5 = 0.029
D B	2	(3·10)/20 = 1.5	(1-1.5)²/1.5 = 0.167

	A	C	
L	9	6	15
D	1	4	5
	10	10	20

A/C	O	E	Sub-group
L A	9	(15·10)/20 = 7.5	(9-7.5)²/7.5 = 0.3
D A	1	(5·10)/20 = 2.5	(1-2.5)²/2.5 = 0.9
L C	6	(15·10)/20 = 7.5	(9-7.5)²/7.5 = 0.3
D C	4	(5·10)/20 = 2.5	(1-2.5)²/2.5 = 0.9

DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

Chi-Squared 3

UND UNIVERSITY OF NORTH DAKOTA

	Total	A	B	C
L	23	9	8	6
D	7	1	2	4

$$\chi^2 = \sum_{i,j} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

$\chi^2_{AB} = 0.019 \cdot 2 + 0.167 \cdot 2 = 0.39$

$\chi^2_{Ac} = 0.3 \cdot 2 + 0.9 \cdot 2 = 2.4$

$df = (2-1) \cdot (2-1) = 1$

$\chi^2_{crit} = 3.841$

Steps

1. State hypothesis
2. Select alpha
3. Fill out table
4. Calculate observed values (O)
5. Calculate expected values (E)
6. Calculate subgroup values
7. **Calculate test statistic (χ^2)**
8. **Calculate degrees of freedom (df)**
9. **Look up test statistic in table**
10. State conclusion

The Table

	0.995	0.99	0.975	0.95	0.9	0.5	0.2	0.1	0.05	0.025	0.02	0.01	0.005	0.002	0.001
1	0.0000397	0.000157	0.000982	0.00393	0.0158	0.455	1.642	2.706	3.841	5.024	5.412	6.635	7.879	9.550	10.828
2	0.0100	0.020	0.051	0.103	0.211	1.386	3.219	4.605	5.991	7.378	7.824	9.210	10.597	12.429	13.816
3	0.072	0.115	0.216	0.352	0.584	2.366	4.642	6.251	7.815	9.348	9.837	11.345	12.838	14.796	16.266
4	0.207	0.297	0.484	0.711	1.064	3.357	5.989	7.779	9.488	11.143	11.668	13.277	14.860	16.924	18.467
5	0.412	0.554	0.831	1.145	1.610	4.351	7.289	9.236	11.070	12.833	13.388	15.086	16.750	18.907	20.515
6	0.676	0.872	1.237	1.635	2.204	5.348	8.558	10.645	12.592	14.449	15.033	16.812	18.548	20.791	22.458
7	0.989	1.239	1.690	2.167	2.833	6.346	9.803	12.017	14.067	16.013	16.622	18.475	20.278	22.601	24.322
8	1.344	1.646	2.180	2.733	3.490	7.344	11.030	13.362	15.507	17.535	18.168	20.090	21.955	24.352	26.124
9	1.735	2.088	2.700	3.325	4.168	8.343	12.242	14.684	16.919	19.023	19.679	21.666	23.589	26.056	27.877
10	2.156	2.558	3.247	3.940	4.865	9.342	13.442	15.987	18.307	20.483	21.161	23.209	25.188	27.722	29.588
11	2.603	3.053	3.816	4.575	5.578	10.341	14.631	17.275	19.675	21.920	22.618	24.725	26.757	29.354	31.264
12	3.074	3.571	4.404	5.226	6.304	11.340	15.812	18.549	21.026	23.337	24.054	26.217	28.300	30.957	32.909
13	3.565	4.107	5.009	5.892	7.042	12.340	16.985	19.812	22.362	24.736	25.472	27.688	29.819	32.535	34.528
14	4.075	4.660	5.629	6.571	7.790	13.339	18.151	21.064	23.685	26.119	26.873	29.141	31.319	34.091	36.123
15	4.601	5.229	6.262	7.261	8.547	14.339	19.311	22.307	24.996	27.488	28.259	30.578	32.801	35.628	37.697

[5]

DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

Chi-Squared 4

UND UNIVERSITY OF NORTH DAKOTA

	Total	A	B	C
L	23	9	8	6
D	7	1	2	4

Steps

1. State hypothesis
2. Select alpha
3. Fill out table
4. Calculate observed values (O)
5. Calculate expected values (E)
6. Calculate subgroup values
7. Calculate test statistic (χ^2)
8. Calculate degrees of freedom (df)
9. Look up test statistic in table
10. **State conclusion**

Conclusion: neither AB nor AC had significant X_2 test statistic, as they were below the critical X_2 test

Looking further: how would the data need to look to get a significant test?

	X	Y	
L	9	1	10
D	1	9	10
	10	10	20

	O	E	sub-group
LX	9	5	$(9-5)^2/5 = 3.2$
DX	1	5	$(1-5)^2/5 = 3.2$
LY	1	5	$(9-5)^2/5 = 3.2$
DY	9	5	$(1-5)^2/5 = 3.2$
			$+ 12.8$ (df=1)

DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

T-tests

UND UNIVERSITY OF NORTH DAKOTA

[6][7][8]

Steps

1. State hypothesis
2. Select alpha & tail
3. Sketch graph
4. Calculate group means ($\bar{x}$)
5. Calculate group sizes (n)
6. Calculate group standard deviations (s^2)
7. Calculate test statistic (t)
8. Calculate degrees of freedom (df)
9. Look up test statistic in table
10. State conclusion

$$df = (n1 + n2) - 2$$

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

$$\sigma = \sqrt{\frac{\sum(X-\mu)^2}{n}}$$

$$s = \sqrt{\frac{\sum(X-\bar{X})^2}{n-1}}$$

DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

T-tests 2

UND UNIVERSITY OF NORTH DAKOTA

ID	Location	Length
1	A	5.375
2	A	4.625
3	A	3.125
4	A	3.125
5	A	1.5
6	A	3.5
7	A	2.75
8	A	3.125
9	A	3.25
10	A	4.375

ID	Location	Length
11	B	5.5
12	B	2.75
13	B	3.25
14	B	4.625
15	B	3.375
16	B	6.375
17	B	2.75
18	B	2
19	B	4.375
20	B	4.375

Steps

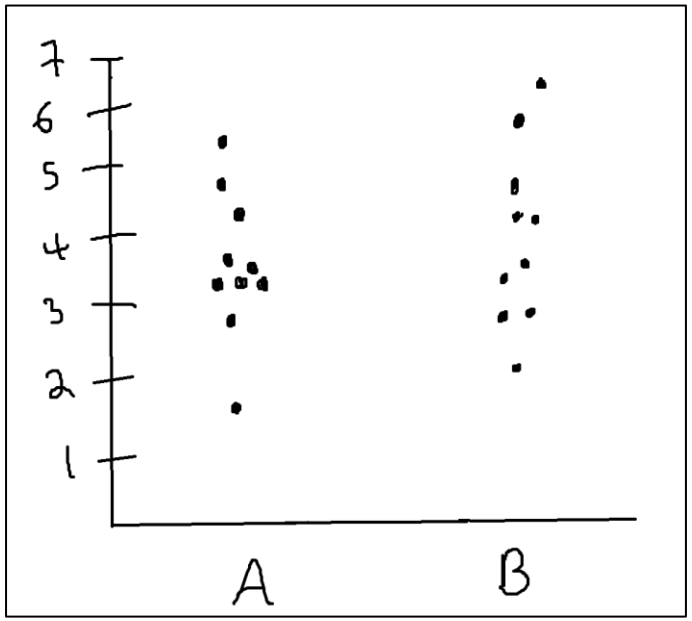
1. **State hypothesis**
2. **Select alpha & tail**
3. **Sketch graph**
4. Calculate group means ($\bar{x}$)
5. Calculate group sizes (n)
6. Calculate group standard deviations (s^2)
7. Calculate test statistic (t)
8. Calculate degrees of freedom (df)
9. Look up test statistic in table
10. State conclusion

H_0: $\mu_A = \mu_B$

H_1: $\mu_A \neq \mu_B$

$\alpha = 0.05$

two tailed



DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

T-tests 3

UND UNIVERSITY OF NORTH DAKOTA

Steps

1. State hypothesis
2. Select alpha & tail
3. Sketch graph
4. **Calculate group means ($\bar{x}$)**
5. **Calculate group sizes (n)**
6. **Calculate group standard deviations (s^2)**
7. **Calculate test statistic (t)**
8. Calculate degrees of freedom (df)
9. Look up test statistic in table
10. State conclusion

	Section A Length	Section B Length		A Xi	X	(Xi-X)^2		B Xi	X	(Xi-X)^2
	5.375	5.5		5.375	3.475	3.61		5.5	3.9375	2.441406
	4.625	2.75		4.625	3.475	1.3225		2.75	3.9375	1.410156
	3.125	3.25		3.125	3.475	0.1225		3.25	3.9375	0.472656
	3.125	4.625		3.125	3.475	0.1225		4.625	3.9375	0.472656
	1.5	3.375		1.5	3.475	3.900625		3.375	3.9375	0.316406
	3.5	6.375		3.5	3.475	0.000625		6.375	3.9375	5.941406
	2.75	2.75		2.75	3.475	0.525625		2.75	3.9375	1.410156
	3.125	2		3.125	3.475	0.1225		2	3.9375	3.753906
	3.25	4.375		3.25	3.475	0.050625		4.375	3.9375	0.191406
	4.375	4.375		4.375	3.475	0.81		4.375	3.9375	0.191406
sum	34.75	39.375			sum	10.5875			sum	16.60156
N	10	10			SD	1.084615			SD	1.358167
mean	3.475	3.9375								

$$s = \sqrt{\frac{\sum(X-\bar{X})^2}{n-1}}$$

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

$$t = \frac{(3.475 - 3.9375)}{\sqrt{\frac{1.085^2}{10} + \frac{1.358^2}{10}}} = \frac{-0.4625}{\sqrt{0.301}} = \frac{-0.46}{0.55} = -0.84$$

DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

T-tests 4

[9]

UND UNIVERSITY OF NORTH DAKOTA.

Steps

1. State hypothesis
2. Select alpha & tail
3. Sketch graph
4. Calculate group means ($\bar{x}$)
5. Calculate group sizes (n)
6. Calculate group standard deviations (s^2)
7. Calculate test statistic (t)
8. Calculate degrees of freedom (df)
9. Look up test statistic in table
10. State conclusion

Conclusion: A and B did not have significantly different mean length, as the t-statistic was smaller than the critical t-statistic

t-test table

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	0.50	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
two-tails	1.00	0.50	0.40	0.30	0.20	0.10	0.05	0.02	0.01	0.002	0.001
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850

Looking further: how would the data need to look to get a significant test?

$\frac{(3.5-3.9)}{0.55}$ → boost signal ~ $\frac{(3.5-5.0)}{0.55} = \frac{-1.5}{0.55} = -2.73$

→ reduce noise ~ $\frac{-0.4}{0.15} = -2.67$

DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

ANOVA

UND UNIVERSITY OF NORTH DAKOTA

[10][11]

Steps

1. State hypothesis and alpha
2. Sketch graph
3. Calculate group means and overall mean
4. Calculate regression sum of squares (SSR)
5. Calculate error sum of squares (SSE)
6. Calculate total sum of squares (SST)
7. Calculate degrees of freedom (dfe)
8. Calculate mean squares of treatment (MS)
9. Calculate mean squares of error (ME)
10. Calculate test statistic (F)
11. Look up test statistic in table
12. State conclusion

ANOVA Table: Formulas

Source	df	SS	MS (Mean Square)	F
Model (between)	$I-1$	$\sum_{i=1}^{I} n_i(\bar{x}_{i\cdot}-\bar{x}_{\cdot\cdot})^2$	$\frac{SSM}{dfm}=\frac{SSM}{I-1}$	$\frac{MSM}{MSE}$
Error (within)	$N-I$	$\sum_{i=1}^{I}(n_i-1)s_i^2$	$\frac{SSE}{dfe}=\frac{SSE}{N-I}$	
Total	$N-1$	$\sum_{i=1}^{I}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_{\cdot\cdot})^2$		

DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

ANOVA 2

UND UNIVERSITY OF NORTH DAKOTA

ID	Location	Length
1	A	5.375
2	A	4.625
3	A	3.125
4	A	3.125
5	A	1.5
6	A	3.5
7	A	2.75
8	A	3.125
9	A	3.25
10	A	4.375

ID	Location	Length
11	B	5.5
12	B	2.75
13	B	3.25
14	B	4.625
15	B	3.375
16	B	6.375
17	B	2.75
18	B	2
19	B	4.375
20	B	4.375

ID	Location	Length
21	C	3.125
22	C	3
23	C	4.25
24	C	4.25
25	C	7.25
26	C	6.75
27	C	3.125
28	C	3.375
29	C	3.125
30	C	2.25

Steps

1. **State hypothesis and alpha**
2. **Sketch graph**
3. Calculate group means and overall mean
4. Calculate regression sum of squares (SSR)
5. Calculate error sum of squares (SSE)
6. Calculate total sum of squares (SST)
7. Calculate degrees of freedom (dfe)
8. Calculate mean squares of treatment (MS)
9. Calculate mean squares of error (ME)
10. Calculate test statistic (F)
11. Look up test statistic in table
12. State conclusion

H_0

$\mu_A = \mu_B = \mu_C$

H_1

$\mu_A \neq \mu_B = \mu_C$

or

$\mu_A = \mu_B \neq \mu_C$

or

$\mu_A \neq \mu_B \neq \mu_C$

$\alpha = 0.05$

DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

ANOVA 2

UND UNIVERSITY OF NORTH DAKOTA.

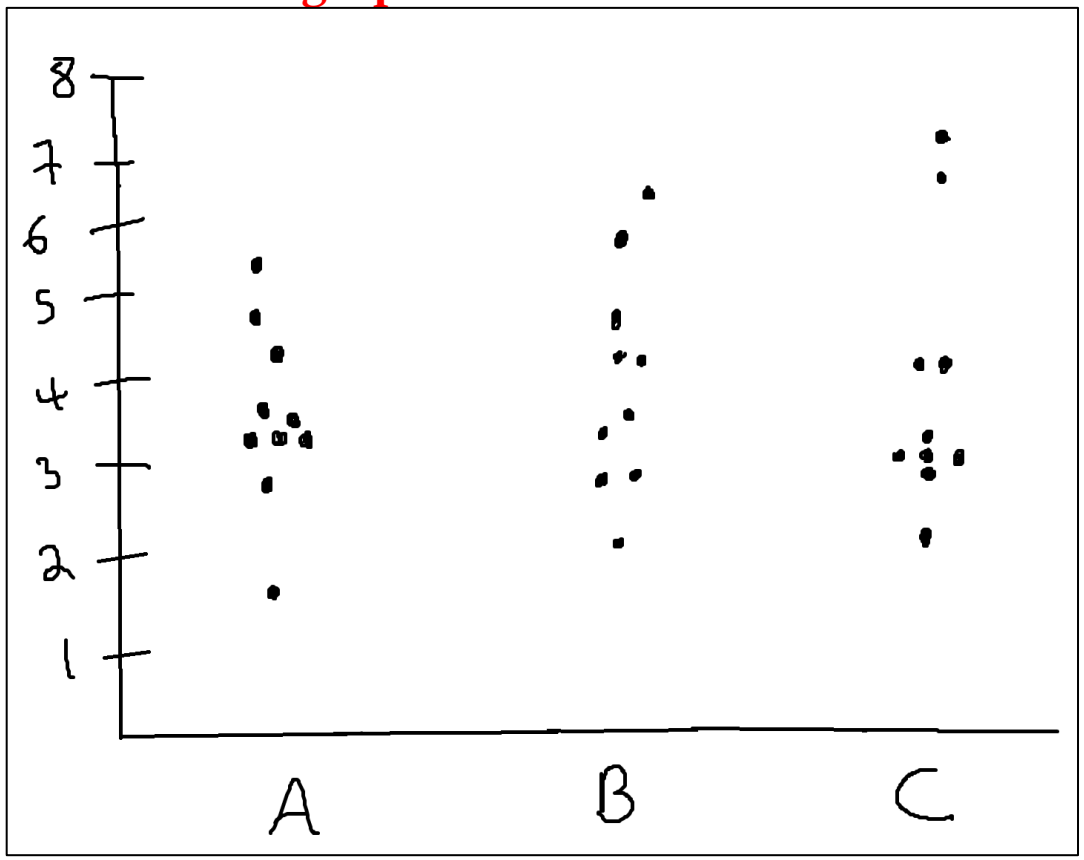
ID	Location	Length
1	A	5.375
2	A	4.625
3	A	3.125
4	A	3.125
5	A	1.5
6	A	3.5
7	A	2.75
8	A	3.125
9	A	3.25
10	A	4.375

ID	Location	Length
11	B	5.5
12	B	2.75
13	B	3.25
14	B	4.625
15	B	3.375
16	B	6.375
17	B	2.75
18	B	2
19	B	4.375
20	B	4.375

ID	Location	Length
21	C	3.125
22	C	3
23	C	4.25
24	C	4.25
25	C	7.25
26	C	6.75
27	C	3.125
28	C	3.375
29	C	3.125
30	C	2.25

Steps

1. State hypothesis and alpha
2. Sketch graph



H_0

$\mu_A = \mu_B = \mu_C$

H_1

$\mu_A \neq \mu_B = \mu_C$

or

$\mu_A = \mu_B \neq \mu_C$

or

$\mu_A \neq \mu_B \neq \mu_C$

$\alpha = 0.05$

DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

ANOVA 3

UNIVERSITY OF NORTH DAKOTA

Steps

1. State hypothesis and alpha
2. Sketch graph
3. **Calculate group means and overall mean**
4. **Calculate regression sum of squares (SSR)**
5. **Calculate error sum of squares (SSE)**
6. **Calculate total sum of squares (SST)**
7. Calculate degrees of freedom (dfe)
8. Calculate mean squares of treatment (MS)
9. Calculate mean squares of error (ME)
10. Calculate test statistic (F)
11. Look up test statistic in table
12. State conclusion

	Section A Length	Section B Length	Section C Length
	5.375	5.5	3.125
	4.625	2.75	3
	3.125	3.25	4.25
	3.125	4.625	4.25
	1.5	3.375	7.25
	3.5	6.375	6.75
	2.75	2.75	3.125
	3.125	2	3.375
	3.25	4.375	3.125
	4.375	4.375	2.25
Group Means	3.475	3.938	4.05
Overall Mean	3.821		

Source	df	SS
Model (between)	$I-1$	$\sum_{i=1}^{I} n_i(\bar{x}_{i.}-\bar{x}_{..})^2$
Error (within)	$N-I$	$\sum_{i=1}^{I}(n_i-1)s_i^2$
Total	$N-1$	$\sum_{i=1}^{I}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_{..})^2$

	SSR		SSE A	SSE B	SSE C		SST
	1.19716		3.61	2.439844	0.855625		54.02253
	0.13689		1.3225	1.411344	1.1025		
	0.52441		0.1225	0.473344	0.04		
Total	1.85846		0.1225	0.471969	0.04		
			3.900625	0.316969	10.24		
			0.000625	5.938969	7.29		
			0.525625	1.411344	0.855625		
			0.1225	3.755844	0.455625		
			0.050625	0.190969	0.855625		
			0.81	0.190969	3.24		
		Sub	10.5875	16.60157	24.975		
		Total	52.16407				

DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

ANOVA 4

UND UNIVERSITY OF NORTH DAKOTA.

Steps

1. State hypothesis and alpha
2. Sketch graph
3. Calculate group means and overall mean
4. Calculate regression sum of squares (SSR)
5. Calculate error sum of squares (SSE)
6. Calculate total sum of squares (SST)
7. **Calculate degrees of freedom (dfe)**
8. **Calculate mean squares of treatment (MS)**
9. **Calculate mean squares of error (ME)**
10. **Calculate test statistic (F)**
11. Look up test statistic in table
12. State conclusion

ANOVA Table				
Source	**SS**	**df**	**Mean Sq**	**F**
Treatment	1.86			
Error	52.16			
Total	54.02			

$df_{Treat} = \#\,groups - 1$

$df_{error} = total\ n - \#\,groups$

$df_{tot} = total\ n - 1$

$MS = SSR / df_{Treat}$

$ME = SSE / df_{error}$

$F = \frac{MS}{ME}$

DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

ANOVA 4

UND UNIVERSITY OF NORTH DAKOTA.

Steps

1. State hypothesis and alpha
2. Sketch graph
3. Calculate group means and overall mean
4. Calculate regression sum of squares (SSR)
5. Calculate error sum of squares (SSE)
6. Calculate total sum of squares (SST)
7. **Calculate degrees of freedom (dfe)**
8. **Calculate mean squares of treatment (MS)**
9. **Calculate mean squares of error (ME)**
10. **Calculate test statistic (F)**
11. Look up test statistic in table
12. State conclusion

ANOVA Table				
Source	**SS**	**df**	**Mean Sq**	**F**
Treatment	1.86	**2**	**0.93**	**0.48**
Error	52.16	**27**	**1.93**	
Total	54.02	**29**		

$df_{Treat} = \#\,groups - 1$

$df_{error} = total\ n - \#\,groups$

$df_{tot} = total\ n - 1$

$MS = SSR / df_{Treat}$

$ME = SSE / df_{error}$

$F = \frac{MS}{ME}$

DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

ANOVA 4

UND UNIVERSITY OF NORTH DAKOTA

[12]

Steps

1. State hypothesis and alpha
2. Sketch graph
3. Calculate group means and overall mean
4. Calculate regression sum of squares (SSR)
5. Calculate error sum of squares (SSE)
6. Calculate total sum of squares (SST)
7. Calculate degrees of freedom (dfe)
8. Calculate mean squares of treatment (MS)
9. Calculate mean squares of error (ME)
10. Calculate test statistic (F)
11. Look up test statistic in table
12. **State conclusion**

Conclusion: No difference between mean length across A, B, and C, as the F-statistic was smaller than the critical F-statistic

$\frac{0.93}{1.93}$ → boost signal ~ $\frac{0.93 \times 10}{1.93} = \frac{9.3}{1.93} = 4.82$

$\frac{0.93}{1.93}$ → reduce noise ~ $\frac{0.93}{1.93 \times 0.1} = \frac{0.93}{0.193} = 4.82$

Looking further: how would the data need to look to get a significant test?

	DF1	α = 0.05			
DF2	1	2	3	4	5
1	161.45	199.5	215.71	224.58	230.16
2	18.513	19	19.164	19.247	19.296
3	10.128	9.5521	9.2766	9.1172	9.0135
4	7.7086	6.9443	6.5914	6.3882	6.2561
5	6.6079	5.7861	5.4095	5.1922	5.0503
6	5.9874	5.1433	4.7571	4.5337	4.3874
7	5.5914	4.7374	4.3468	4.1203	3.9715
8	5.3177	4.459	4.0662	3.8379	3.6875
9	5.1174	4.2565	3.8625	3.6331	3.4817
10	4.9646	4.1028	3.7083	3.478	3.3258
11	4.8443	3.9823	3.5874	3.3567	3.2039
12	4.7472	3.8853	3.4903	3.2592	3.1059
13	4.6672	3.8056	3.4105	3.1791	3.0254
14	4.6001	3.7389	3.3439	3.1122	2.9582
15	4.5431	3.6823	3.2874	3.0556	2.9013
16	4.494	3.6337	3.2389	3.0069	2.8524
17	4.4513	3.5915	3.1968	2.9647	2.81
18	4.4139	3.5546	3.1599	2.9277	2.7729
19	4.3807	3.5219	3.1274	2.8951	2.7401
20	4.3512	3.4928	3.0984	2.8661	2.7109
21	4.3248	3.4668	3.0725	2.8401	2.6848
22	4.3009	3.4434	3.0491	2.8167	2.6613
23	4.2793	3.4221	3.028	2.7955	2.64
24	4.2597	3.4028	3.0088	2.7763	2.6207
25	4.2417	3.3852	2.9912	2.7587	2.603
26	4.2252	3.369	2.9752	2.7426	2.5868
27	4.21	3.3541	2.9604	2.7278	2.5719
28	4.196	3.3404	2.9467	2.7141	2.5581
29	4.183	3.3277	2.934	2.7014	2.5454
30	4.1709	3.3158	2.9223	2.6896	2.5336

DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

Correlation

UND UNIVERSITY OF NORTH DAKOTA.

[13][14]

Steps

1. State hypothesis and tail
2. Sketch graph
3. Calculate means of X and Y
4. Calculate the difference between means
5. Calculate other variables
6. Calculate correlation coefficient (r)
7. Calculate degrees of freedom
8. Look up coefficient in table
9. State conclusion

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

Where,

r = Pearson Correlation Coefficient

x_i = x variable samples

y_i = y variable sample

$\bar{x}$ = mean of values in x variable

$\bar{y}$ =mean of values in y variable

DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

Correlation 2

UND UNIVERSITY OF NORTH DAKOTA.

ID	Length	Width
1	5.375	0.5
2	4.625	0.375
3	3.125	0.875
4	3.125	0.25
5	1.5	0.75
6	3.5	0.625
7	2.75	0.5
8	3.125	0.5
9	3.25	0.625
10	4.375	0.5

ID	Length	Width
11	5.5	0.75
12	2.75	0.5
13	3.25	0.625
14	4.625	0.5
15	3.375	0.5
16	6.375	0.5
17	2.75	0.5
18	2	0.625
19	4.375	0.875
20	4.375	0.5

Steps

1. **State hypothesis and tail**
2. **Sketch graph**
3. Calculate means of X and Y
4. Calculate the difference between means
5. Calculate other variables
6. Calculate correlation coefficient (r)
7. Calculate degrees of freedom
8. Look up coefficient in table
9. State conclusion

Ho
NO significant correlation

Hi
Significant correlation

two tailed

DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

Correlation 2

UND UNIVERSITY OF NORTH DAKOTA

ID	Length	Width
1	5.375	0.5
2	4.625	0.375
3	3.125	0.875
4	3.125	0.25
5	1.5	0.75
6	3.5	0.625
7	2.75	0.5
8	3.125	0.5
9	3.25	0.625
10	4.375	0.5

ID	Length	Width
11	5.5	0.75
12	2.75	0.5
13	3.25	0.625
14	4.625	0.5
15	3.375	0.5
16	6.375	0.5
17	2.75	0.5
18	2	0.625
19	4.375	0.875
20	4.375	0.5

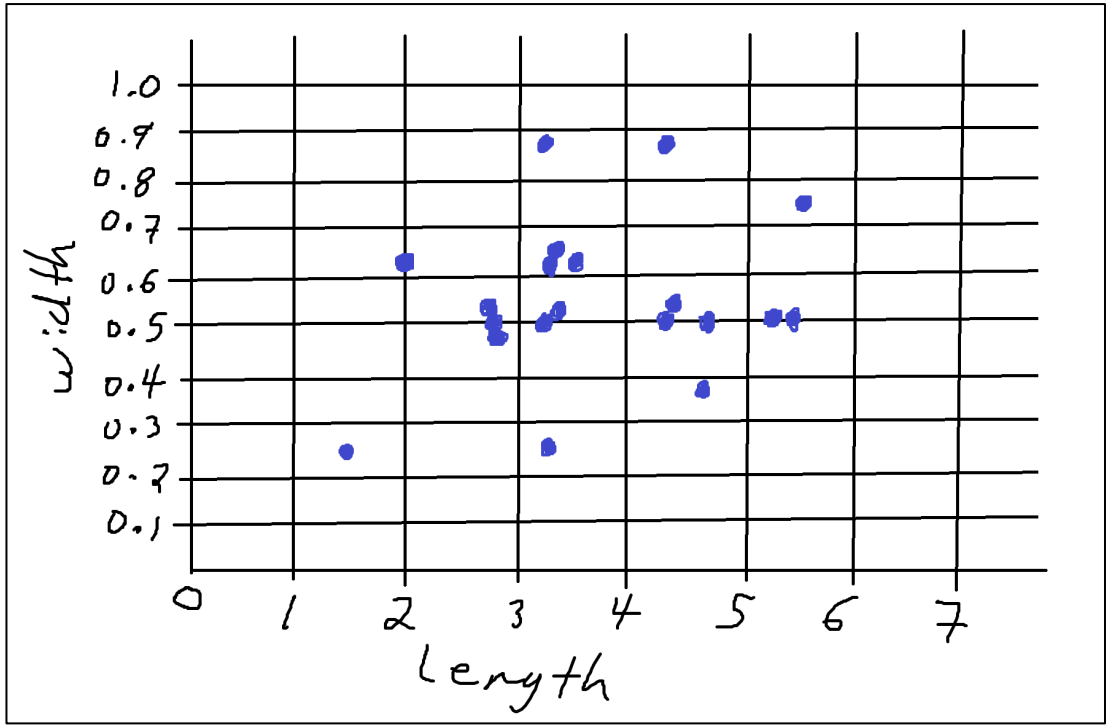
Steps

1. **State hypothesis and tail**
2. **Sketch graph**
3. Calculate means of X and Y

H_0
NO significant correlation

H_1
Significant correlation

two tailed



DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

Correlation 3

UND UNIVERSITY OF NORTH DAKOTA.

Steps

1. State hypothesis and tail
2. Sketch graph
3. **Calculate means of X and Y**
4. **Calculate the difference between means**
5. **Calculate other variables**
6. **Calculate correlation coefficient (r)**
7. Calculate degrees of freedom
8. Look up coefficient in table
9. State conclusion

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

ID	Length (x)	Width (y)	x-xm	y-ym	(x-xm)*(y-ym)	(x-xm)^2	(y-ym)^2
1	5.375	0.5	1.665	-0.07	-0.11655	2.772225	0.0049
2	4.625	0.375	0.915	-0.195	-0.178425	0.837225	0.038025
3	3.125	0.875	-0.585	0.305	-0.178425	0.342225	0.093025
4	3.125	0.25	-0.585	-0.32	0.1872	0.342225	0.1024
5	1.5	0.75	-2.21	0.18	-0.3978	4.8841	0.0324
6	3.5	0.625	-0.21	0.055	-0.01155	0.0441	0.003025
7	2.75	0.5	-0.96	-0.07	0.0672	0.9216	0.0049
8	3.125	0.5	-0.585	-0.07	0.04095	0.342225	0.0049
9	3.25	0.625	-0.46	0.055	-0.0253	0.2116	0.003025
10	4.375	0.5	0.665	-0.07	-0.04655	0.442225	0.0049
11	5.5	0.75	1.79	0.18	0.3222	3.2041	0.0324
12	2.75	0.5	-0.96	-0.07	0.0672	0.9216	0.0049
13	3.25	0.625	-0.46	0.055	-0.0253	0.2116	0.003025
14	4.625	0.5	0.915	-0.07	-0.06405	0.837225	0.0049
15	3.375	0.5	-0.335	-0.07	0.02345	0.112225	0.0049
16	6.375	0.5	2.665	-0.07	-0.18655	7.102225	0.0049
17	2.75	0.5	-0.96	-0.07	0.0672	0.9216	0.0049
18	2	0.625	-1.71	0.055	-0.09405	2.9241	0.003025
19	4.375	0.875	0.665	0.305	0.202825	0.442225	0.093025
20	4.375	0.5	0.665	-0.07	-0.04655	0.442225	0.0049
sum	74.125	11.375		sum	-0.392875	28.25888	0.452375
n	20	20					
mean	3.70625	0.56875					

$r = -0.39 / \sqrt{(28.26 \cdot 0.45)} = -0.39 / 3.57 = -0.11$

DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

Correlation 4

[15]

UND UNIVERSITY OF NORTH DAKOTA.

Steps

1. State hypothesis and tail
2. Sketch graph
3. Calculate means of X and Y
4. Calculate the difference between means
5. Calculate other variables
6. Calculate correlation coefficient (r)
7. **Calculate degrees of freedom**
8. **Look up coefficient in table**
9. **State conclusion**

$df = \text{total } n - 2 = 18$

Conclusion: Correlation not significant, as r-statistic is smaller than critical r-statistic.

Pearson's Correlation Table

df \ α	0.2	0.1	0.05	0.02	0.01	0.001
1	0.951057	0.987688	0.996917	0.999507	0.999877	0.999999
2	0.800000	0.900000	0.950000	0.980000	0.990000	0.999000
3	0.687049	0.805384	0.878339	0.934333	0.958735	0.991139
4	0.608400	0.729299	0.811401	0.882194	0.917200	0.974068
5	0.550863	0.669439	0.754492	0.832874	0.874526	0.950883
6	0.506727	0.621489	0.706734	0.788720	0.834342	0.924904
7	0.471589	0.582206	0.666384	0.749776	0.797681	0.898260
8	0.442796	0.549357	0.631897	0.715459	0.764592	0.872115
9	0.418662	0.521404	0.602069	0.685095	0.734786	0.847047
10	0.398062	0.497265	0.575983	0.658070	0.707888	0.823305
11	0.380216	0.476156	0.552943	0.633863	0.683528	0.800962
12	0.364562	0.457500	0.532413	0.612047	0.661376	0.779998
13	0.350688	0.440861	0.513977	0.592270	0.641145	0.760351
14	0.338282	0.425902	0.497309	0.574245	0.622591	0.741934
15	0.327101	0.412360	0.482146	0.557737	0.605506	0.724657
16	0.316958	0.400027	0.468277	0.542548	0.589714	0.708429
17	0.307702	0.388733	0.455531	0.528517	0.575067	0.693163
18	0.299210	0.378341	0.443763	0.515505	0.561435	0.678781
19	0.291384	0.368737	0.432858	0.503397	0.548711	0.665208
20	0.284140	0.359827	0.422714	0.492094	0.536800	0.652378
21	0.277411	0.351531	0.413247	0.481512	0.525620	0.640230
22	0.271137	0.343783	0.404386	0.471579	0.515101	0.628710
23	0.265270	0.336524	0.396070	0.462231	0.505182	0.617768
24	0.259768	0.329705	0.388244	0.453413	0.495808	0.607360
25	0.254594	0.323283	0.380863	0.445078	0.486932	0.597446
26	0.249717	0.317223	0.373886	0.437184	0.478511	0.587988
27	0.245110	0.311490	0.367278	0.429693	0.470509	0.578956
28	0.240749	0.306057	0.361007	0.422572	0.462892	0.570317
29	0.236612	0.300898	0.355046	0.415792	0.455631	0.562047
30	0.232681	0.295991	0.349370	0.409327	0.448699	0.554119

DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

Regression

UND UNIVERSITY OF NORTH DAKOTA

[16][17]

Steps

1. State hypothesis
2. Sketch graph
3. Calculate X*Y, X^2, and Y^2
4. Calculate sums
5. Calculate intercept (b_0)
6. Calculate slope (b_1)
7. Fill out equation
8. Add line to graph
9. State conclusion

$$b_0 = \frac{(\Sigma y)(\Sigma x^2) - (\Sigma x)(\Sigma xy)}{n(\Sigma x^2) - (\Sigma x)^2}$$

$$b_1 = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{n(\Sigma x^2) - (\Sigma x)^2}$$

Regression 2

ID	Length	Width
1	5.375	0.5
2	4.625	0.375
3	3.125	0.875
4	3.125	0.25
5	1.5	0.75
6	3.5	0.625
7	2.75	0.5
8	3.125	0.5
9	3.25	0.625
10	4.375	0.5

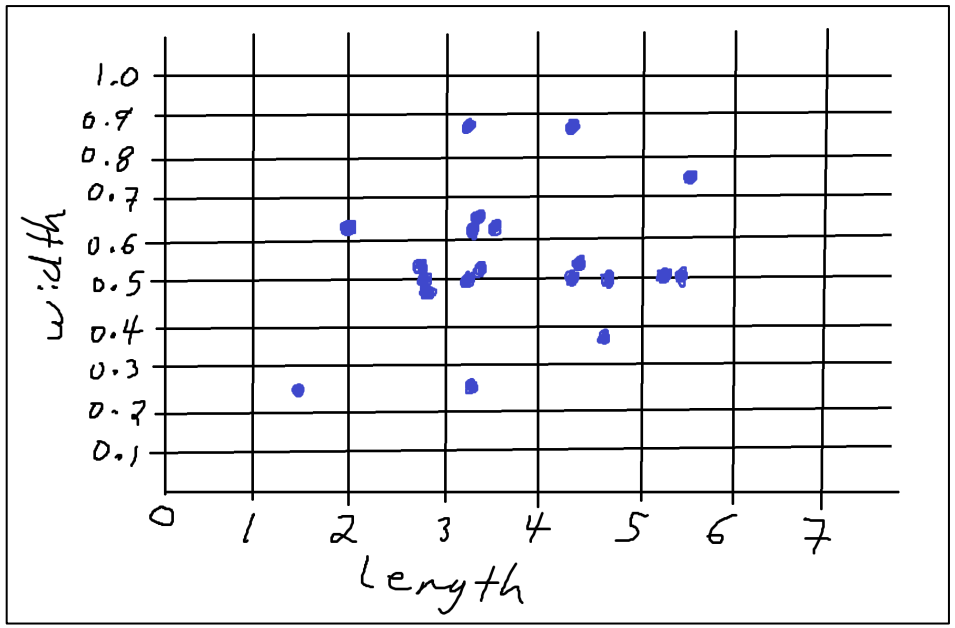
ID	Length	Width
11	5.5	0.75
12	2.75	0.5
13	3.25	0.625
14	4.625	0.5
15	3.375	0.5
16	6.375	0.5
17	2.75	0.5
18	2	0.625
19	4.375	0.875
20	4.375	0.5

Steps

1. **State hypothesis**
2. **Sketch graph**
3. Calculate X*Y, X^2, and Y^2
4. Calculate sums
5. Calculate intercept (b_0)
6. Calculate slope (b_1)
7. Fill out equation
8. Add line to graph
9. State conclusion

H_0
NO significant relationship

H_1
significant relationship



Regression 3

Steps

1. State hypothesis
2. Sketch graph
3. **Calculate X*Y, X^2, and Y^2**
4. **Calculate sums**
5. **Calculate intercept (b_0)**
6. **Calculate slope (b_1)**
7. **Fill out equation**
8. Add line to graph
9. State conclusion

	Length (x)	Width (y)	x*y	x^2	y^2
	5.375	0.5	2.6875	28.89063	0.25
	4.625	0.375	1.734375	21.39063	0.140625
	3.125	0.875	2.734375	9.765625	0.765625
	3.125	0.25	0.78125	9.765625	0.0625
	1.5	0.75	1.125	2.25	0.5625
	3.5	0.625	2.1875	12.25	0.390625
	2.75	0.5	1.375	7.5625	0.25
	3.125	0.5	1.5625	9.765625	0.25
	3.25	0.625	2.03125	10.5625	0.390625
	4.375	0.5	2.1875	19.14063	0.25
	5.5	0.75	4.125	30.25	0.5625
	2.75	0.5	1.375	7.5625	0.25
	3.25	0.625	2.03125	10.5625	0.390625
	4.625	0.5	2.3125	21.39063	0.25
	3.375	0.5	1.6875	11.39063	0.25
	6.375	0.5	3.1875	40.64063	0.25
	2.75	0.5	1.375	7.5625	0.25
	2	0.625	1.25	4	0.390625
	4.375	0.875	3.828125	19.14063	0.765625
	4.375	0.5	2.1875	19.14063	0.25
sum	74.125	11.375	41.76563	302.9844	6.921875

$$b_0 = \frac{(\Sigma y)(\Sigma x^2) - (\Sigma x)(\Sigma xy)}{n(\Sigma x^2) - (\Sigma x)^2}$$

$$b_1 = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{n(\Sigma x^2) - (\Sigma x)^2}$$

$$b_0 = \frac{(11.38)(302.98) - (74.13)(41.77)}{20(302.98) - (74.13)^2} = \frac{351.50}{564.34} = 0.62$$

$$b_1 = \frac{20(41.77) - (74.13)(11.38)}{20(302.98) - (74.13)^2} = \frac{-8.20}{564.34} = -0.01$$

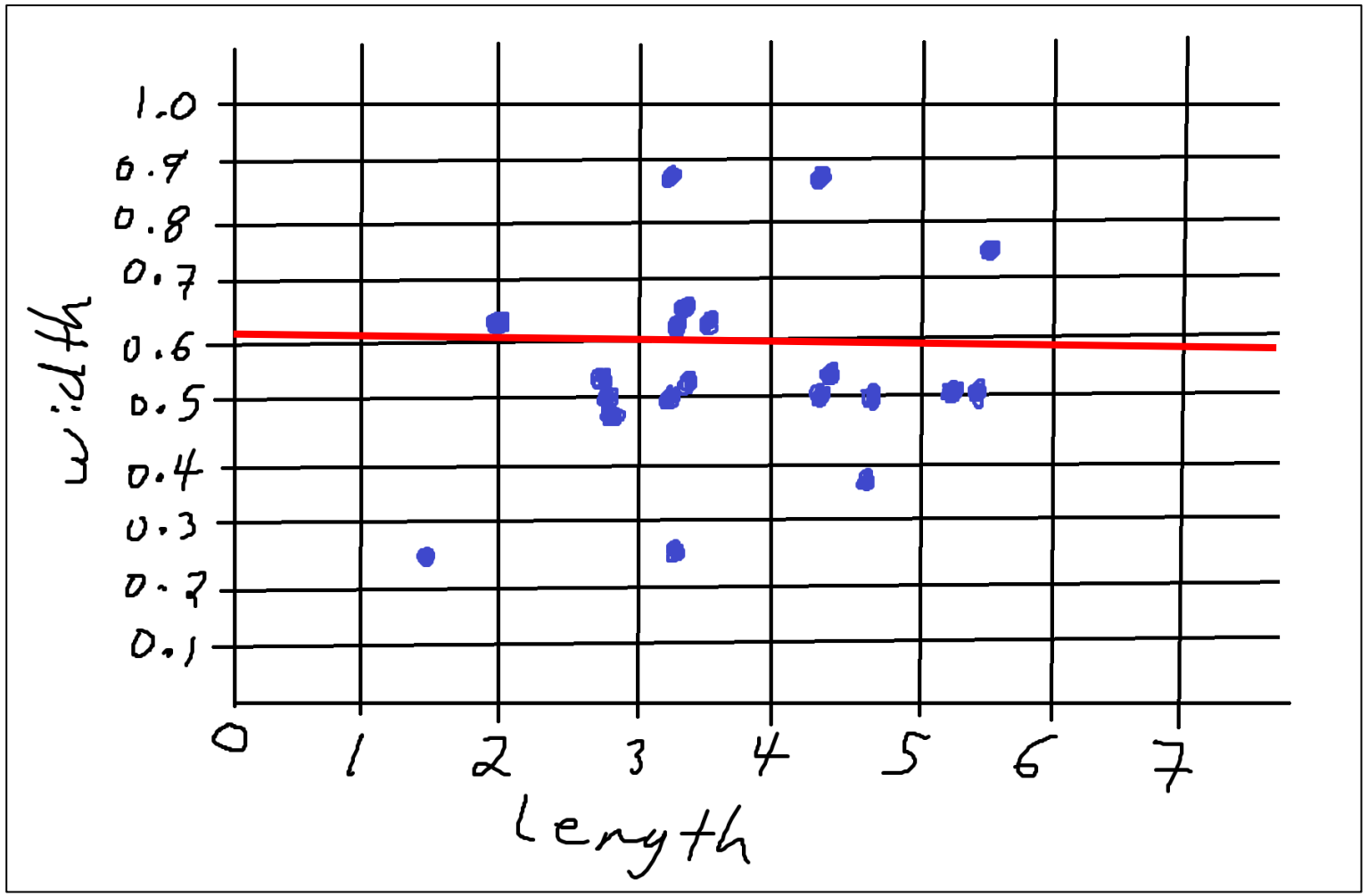
$$y = b_0 + b_1 \cdot x$$

$$\text{width} = 0.62 - 0.01 \cdot \text{Length}$$

Regression 4

Steps

1. State hypothesis
2. Sketch graph
3. Calculate $X*Y$, X^2, and Y^2
4. Calculate sums
5. Calculate intercept (b_0)
6. Calculate slope (b_1)
7. Fill out equation
8. **Add line to graph**
9. **State conclusion**



DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

Comparison

UND
UNIVERSITY OF NORTH DAKOTA.

Method		
Chi-Square	0.39 2.4	
T-test	-0.84	
ANOVA	0.48	
Correlation	-0.11	
Regression	0.62 -0.01	

DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

Comparison

UND
UNIVERSITY OF NORTH DAKOTA.

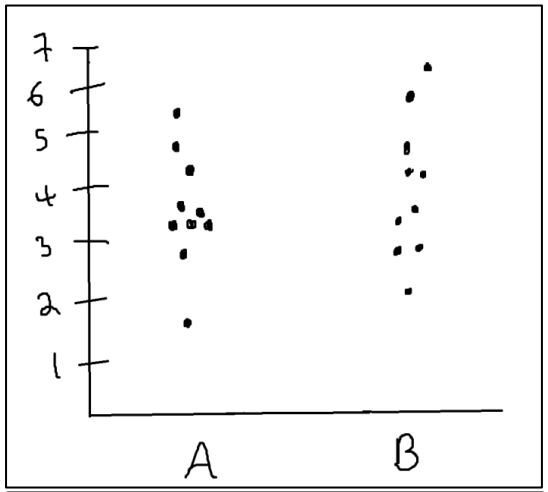
Method		
Chi-Square	0.39 2.4	0.39216 2.4
T-test	-0.84	-0.84146
ANOVA	0.48	0.4807
Correlation	-0.11	-0.1099128
Regression	0.62 -0.01	0.62029 -0.01391

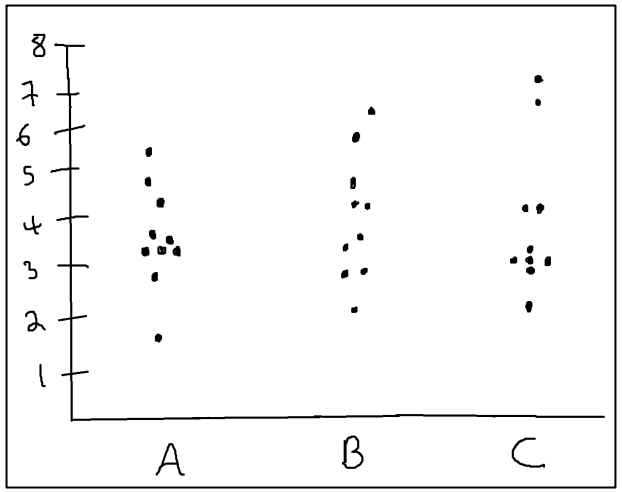
```
#Chi-Square
M1 <- as.table(rbind(c(9,8), c(1,2)))
dimnames(M1) <-list(color=c("L","D"),location=c("A","B"))
chisq.test(M1,correct=F)

M2 <- as.table(rbind(c(9,6), c(1,4)))
dimnames(M2) <-list(color=c("L","D"), location=c("A","C"))
chisq.test(M2,correct=F)

#T-test
t.test(woodchips$Length[1:10],woodchips$Length[11:20])

#ANOVA
summary(lm(woodchips$Length~woodchips$Section))

#Correlation/Regression
cor(woodchips$Length[1:20], woodchips$Width[1:20])
summary(lm(woodchips$Width[1:20]~woodchips$Length[1:20]))
```

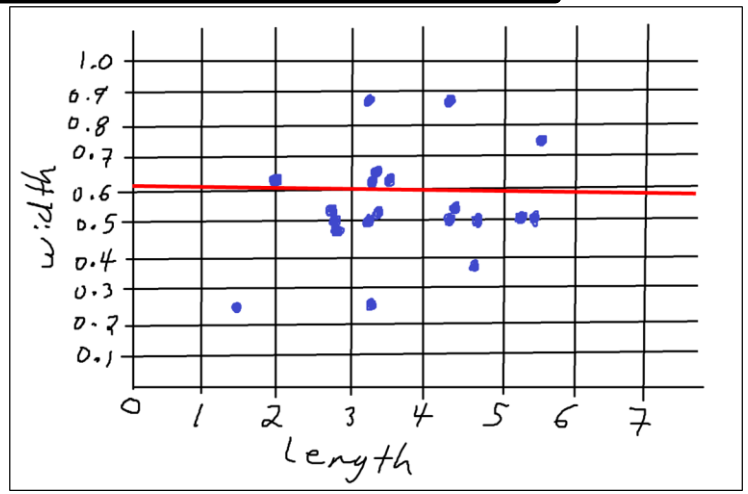
DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

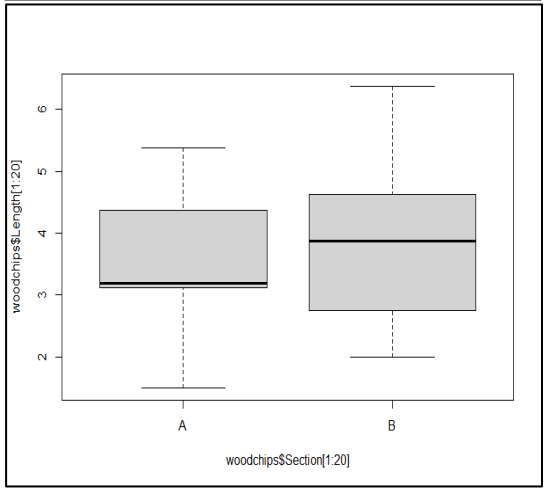
Comparison 2

UNIVERSITY OF NORTH DAKOTA

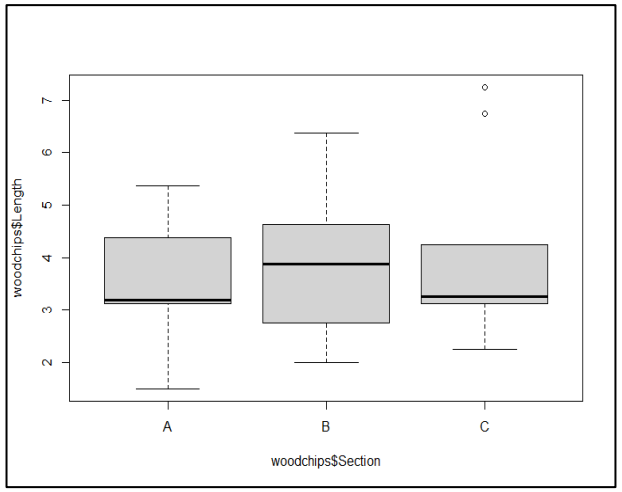




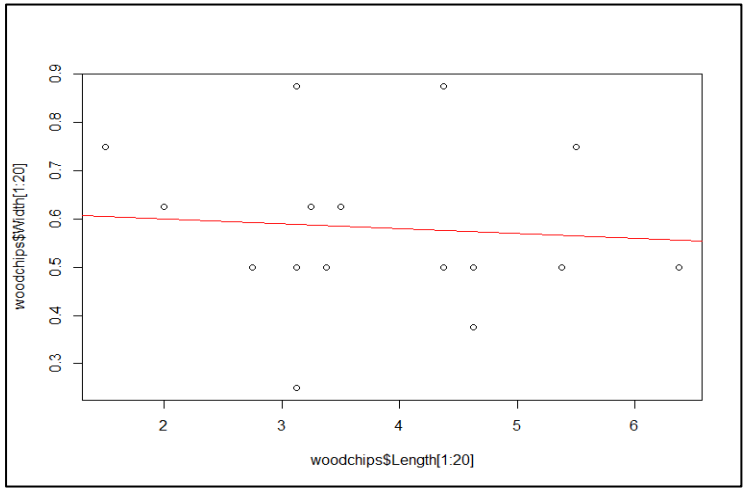




boxplot(woodchips$Length[1:20]~
woodchips$Section[1:20])



boxplot(woodchips$Length~
woodchips$Section)

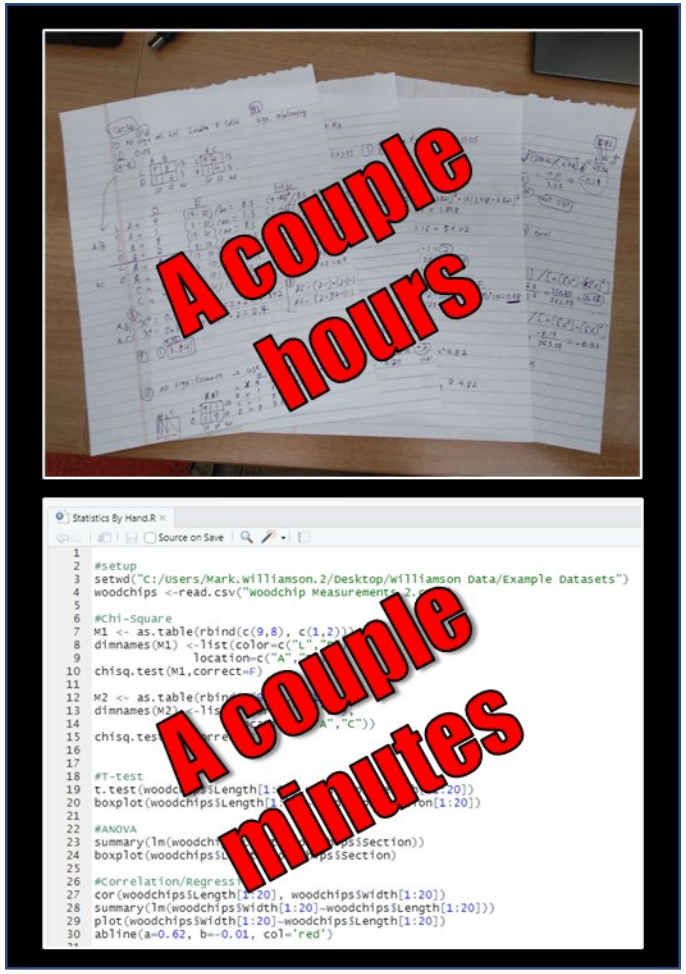


plot(woodchips$Width[1:20]~woodchips$Length[1:20])
abline(a=0.62, b=-0.01, col='red')

DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

Conclusions

- Running classic statistical tests by hand is eminently possible
- Helps tease open the black box of raw numbers to computer output
- Gets it under your nails, as it were
- Feel free to go crazy with more advanced examples (ex. multiple regression)
- Normally, however, just use computers



Please take the post-test and survey:

Post-test: https://und.qualtrics.com/jfe/form/SV_0lElHGTIGSzy20K

Survey: https://und.qualtrics.com/jfe/form/SV_72NSM2UlBWZuBpk

DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

References

UND UNIVERSITY OF NORTH DAKOTA.

[1] https://itfeature.com/statistics/a-short-history-of-statistics

[2] https://image2.slideserve.com/4017969/history-of-statistics-n.jpg

[3] https://www.slideshare.net/superboinkjeni/brief-history-of-statistics-and-its-contributor

[4] https://statsandr.com/blog/chi-square-test-of-independence-by-hand/

[5] https://www.mathsisfun.com/data/chi-square-table.html

[6] http://www.mathandstatistics.com/learn-stats/hypothesis-testing/independent-samples-t-test-by-hand

[7] https://www.educba.com/t-test-formula/

[8] https://byjus.com/standard-deviation-formula/

[9] https://byjus.com/maths/t-test-table/

[10] https://www.statology.org/one-way-anova-by-hand/

[11] https://www.slideserve.com/caesar/chapter-14-one-way-analysis-of-variance-anova

[12] https://www.statology.org/f-distribution-table/

[13] https://www.statology.org/correlation-coefficient-by-hand/

[14] https://www.analyticsvidhya.com/blog/2021/01/beginners-guide-to-pearsons-correlation-coefficient/

[15] http://www.real-statistics.com/statistics-tables/pearsons-correlation-table/

[16] https://www.statology.org/linear-regression-by-hand/

[17] https://owlcation.com/stem/How-to-Create-a-Simple-Linear-Regression-Equation

DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

UND
UNIVERSITY OF NORTH DAKOTA.

Acknowledgements

- The DaCCoTA is supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award Number U54GM128729.
- For the labs that use the Biostatistics, Epidemiology, and Research Design Core in any way, including this Module, please acknowledge us for publications. ***"Research reported in this publication was supported by DaCCoTA (the National Institute of General Medical Sciences of the National Institutes of Health under Award Number U54GM128729)".***

DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY