Surviving statistics 101

mean ($\bar{x}$): $\frac{\Sigma x_n}{n}$ where x_n are $x_1, x_2, x_3 .... x_n$

sum of squares (SS): $\Sigma(x_n-\bar{x})^2$ or $\Sigma x_n^2 - \frac{(\Sigma x_n)^2}{n}$

$\Sigma(x_n-\bar{x})^2$: difference of every observation from the mean, then square, and finally sum for all observations

Σx_n^2: square each observation then sum

$(\Sigma x_n)^2$: sum the observations then square

variance (s^2): $\frac{SS}{n-1}$ (*n* if population; *n-1* if sample)

standard deviation (s): $\sqrt{\frac{SS}{n-1}}$ or $\sqrt{s^2}$

standard error of the mean (SEM): $\frac{s}{\sqrt{n}}$ or $\sqrt{\frac{s^2}{n}}$

confidence interval (CI): $\bar{x} \pm [(t_{0.05,\, n-1})(\text{SEM})]$

$t_{0.05,\, n-1}$: Table A.2. (two tailed) where 0.05 is α (x axis) and n-1 is the degrees of freedom (df, y axis)

T test for independent samples

H_o: $mean_1 = mean_2$: there is no significant difference between the groups
H_A: $mean_1 \neq mean_2$: there is a significant difference between the groups

$T_{table} > T_{data}$: accept H_o
$T_{table} < T_{data}$: reject H_o

$$t_{data} = \left| \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \right|$$

where, for example, $\bar{x}_1$ is the mean for group 1, s^2_1 is the variance for group1, and n_1 is the sample size for group 1

T_{table}: $t_{0.05,\ n-1}$

Table A.2. (two tailed) where 0.05 is α (x axis) and n-1 is the degrees of freedom (df, y axis) from the group with smallest sample size

a more powerful T test when assuming equal variances

H_o: $mean_1 = mean_2$: there is no significant difference between the groups
H_A: $mean_1 \neq mean_2$: there is a significant difference between the groups

$T_{table} > T_{data}$: accept H_o
$T_{table} < T_{data}$: reject H_o

$$t_{data} = \left| \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\frac{SS_1+SS_2}{df_1+df_2}}{n_1} + \frac{\frac{SS_1+SS_2}{df_1+df_2}}{n_2}}} \right|$$

where, for example, $\bar{x}_1$ is the mean for group 1, SS_1 is the sum of squares for group 1, df_1 is the degrees of freedom for group 1 (n-1), and n_1 is the sample size for group 1

T_{table}: $t_{0.05,\ n_t-2}$

Table A.2. (two tailed) where 0.05 is α (x axis) and n_t-2 is the total sample size (n_1+n_2)-2 (y axis)

Paired T test for dependent samples (e.g., before vs. after)

H_o: $mean_1 - mean_2 = 0$: there is no significant difference between the groups
H_A: $mean_1 - mean_2 \neq 0$: there is a significant difference between the groups

$T_{table} > T_{data}$: accept H_o
$T_{table} < T_{data}$: reject H_o

$$t_{data} = \frac{\overline{X}\,(x_a - x_b)}{SEM(x_a - x_b)}$$

where $\overline{x}(x_a-x_b)$ is the mean value for the difference of group a (beginning) from group b (end) (every pair of values is subtracted and then the mean difference is obtained for all pairs), and $SEM\,(x_a-x_b)_1$ is the standard error of the mean for the difference of group a (beginning) from group b (end).

T_{table}: $t_{0.05,\ n-1}$

Table A.2. (two tailed) where 0.05 is α (x axis) and n-1 is the degrees of freedom (df, y axis) for the total number of pairs-1.

Single sample T test to test your result vs. that of a population

H_o: mean (from your study) = mean from the population (the population parameters must be known)
H_A: mean (from your study) Ä mean from the population (the population parameters must be known)

$$t_{data} = \frac{\overline{X} - \mu}{SEM_x}$$

where $\overline{x}$ is the mean value for your study, μ is the mean value for the population, and SEM_x is the standard error of the mean four your study

T_{table}: $t_{0.05,\, n-1}$

Table A.2. (two tailed) where 0.05 is α (x axis) and n-1 is the degrees of freedom (your sample size -1).

$T_{table} > T_{data}$: accept H_o
$T_{table} < T_{data}$: reject H_o

---

one way ANOVA (analysis of variance)

H_o: $mean_1 = mean_2 = mean_3 = mean_n$:no significant difference between groups
H_A: $mean_1$ Ä $mean_2$ Ä $mean_3$ Ä $mean_n$: at least one group is a significant different

	$Group_A$	$Group_B$	$Group_C$
	A_1	B_1	C_1
	A_2	B_2	C_2
	A_3	B_3	C_3
	.	.	.
	.	.	.
	.	.	.
	.	.	.
	A_n	B_n	C_n
mean for each group	$\overline{x}_A$	$\overline{x}_B$	$\overline{x}_C$
sample size for each group	n_A	n_B	n_C
sum of x for each group	Σx_A	Σx_B	Σx_C
sum of x^2 for each group	Σx_A^2	Σx_B^2	Σx_C^2
sum of x then squared for each group	$(\Sigma x_A)^2$	$(\Sigma x_B)^2$	$(\Sigma x_C)^2$
sum of x then squared and then divided by the sample size for each group	$\frac{(\Sigma x_A)^2}{n_A}$	$\frac{(\Sigma x_B)^2}{n_B}$	$\frac{(\Sigma x_C)^2}{n_C}$

one way ANOVA (analysis of variance)

to calculate

1) $n_t = (n_A + n_B + n_C)$

2) $x_t = (\Sigma x_A + \Sigma x_B + \Sigma x_C)$

3) $x_t^2 = (\Sigma x_A^2 + \Sigma x_B^2 + \Sigma x_C^2)$

4) $(x_t)^2 = [(\Sigma x_A) + (\Sigma x_B) + (\Sigma x_C)]^2$

5) $\sum \frac{(\Sigma x_n)^2}{n_n} = \left[\frac{(\Sigma x_A)^2}{n_A} + \frac{(\Sigma x_B)^2}{n_B} + \frac{(\Sigma x_C)^2}{n_C} \right]$

6) $\frac{(x_t)^2}{n_t}$

7) $SS_{total} = x_t^2 - \frac{(x_t)^2}{n_t}$

8) $SS_{treatment} = \sum \frac{(\Sigma x_n)^2}{n_n} - \frac{(x_t)^2}{n_t}$

9) $SS_{error} = SS_{total} - SS_{treatment}$

10) $df_{treatment}$ = number of groups -1

11) $df_{total} = N_t - 1$

12) $df_{error} = df_{total} - df_{treatment}$

$F_{table}: F_{df_{treatment},\, df_{error}}$

Table A.6. the table is already at α=0.05 and the $df_{treatments}$ (x axis) and df_{error} (y axis)

ANOVA table

Source of variation	Sum of Squares		Mean squares (MS)	F_{data}
Treatment (between groups)	$SS_{treatment}$	$df_{treatment}$	$\frac{SS_{treatment}}{df_{treatment}}$	$\frac{MS_{treatment}}{MS_{error}}$
Error (within groups)	SS_{error}	df_{error}	$\frac{SS_{error}}{df_{error}}$	
Total	SS_{total}	df_{total}		

$F_{table} > F_{data}$: accept H_o

$F_{table} < F_{data}$: reject H_o

Tukey's Multiple Comparisons test

you need the MS_{error} from the ANOVA for the multiple group comparison

Critical Value (CV): $q\sqrt{\dfrac{MS_{error}}{\left(\dfrac{2(n_A \infty n_B)}{n_A + n_B}\right)}}$

where MS_{error} is obtained from the ANOVA table, n_A is the sample size for group A, n_B is the sample size for group B, and q is the studentized range value obtained form Table A.7.

$$\underbrace{q_{groups,df_{error}}}$$

Table A.7. the table is already at α=0.05 and data are needed from the ANOVA table: groups represents the total number of treatments (x axis) and df_{error} (y axis)

How to compare groups A, B, and C:

A vs. B: if $(\bar{x}_A - \bar{x}_B) > CV$ then they are significantly different
A vs. C: if $(\bar{x}_A - \bar{x}_C) > CV$ then they are significantly different
B vs. C: if $(\bar{x}_B - \bar{x}_C) > CV$ then they are significantly different

---

Linear Regression (least squares regression)

linear equation = $a + bx$

where a is intercept at the y axis and b is the slope or rate of change per unit of x axis

X	Y
X_1	Y_1
X_2	Y_2
.	.
.	.
X_n	Y_n

$$\text{slope } (b) = \frac{\Sigma xy - \dfrac{\Sigma x \Sigma y}{n}}{\Sigma x^2 - \dfrac{(\Sigma x)^2}{n}}$$

$$\text{intercept} = \bar{y} - b\bar{x}$$

Regression ANOVA

H_o: b=0: no significant difference from a slope of zero
H_A: bÄ0: significantly different from a slope of zero

to calculate

1) n = number of x-y pairs
2) Σx
3) Σy
4) Σx^2
5) Σy^2
6) $(\Sigma x)^2$
7) $(\Sigma y)^2$
8) Σxy
9) $\Sigma x \Sigma y$
10) $SS_{total} = \Sigma y^2 - \frac{(\Sigma y)^2}{n}$
11) $SS_{regression} = b \left(\Sigma xy - \frac{\Sigma x \Sigma y}{n} \right)$
12) $SS_{error} = SS_{total} - SS_{regresssion}$
13) $df_{regression} = 1$
14) $df_{total} = n - 1$
15) $df_{error} = df_{total} - df_{regression}$

ANOVA table

Source of variation	Sum of Squares		Mean squares (MS)	F_{data}
Regression	$SS_{regression}$	$df_{regression}$	$\frac{SS_{regression}}{df_{regression}}$	$\frac{MS_{regression}}{MS_{error}}$
Error	SS_{error}	df_{error}	$\frac{SS_{error}}{df_{error}}$	
Total	SS_{total}	df_{total}		

F_{table}: $F_{df_{regression},\, df_{error}}$

Table A.6. the table is already at α=0.05 and the $df_{regression}$ (x axis) and df_{error} (y axis)

$F_{table} > F_{data}$: accept H_o
$F_{table} < F_{data}$: reject H_o

Calculating confidence interval of the slope

$$S_b = \sqrt{\frac{MS_{error}}{\Sigma x^2 - \frac{(\Sigma x)^2}{n}}}$$

$$\beta: b \pm [(t_{0.05,\, n-2})(s_b)]$$

Table A.2. (two tailed) where 0.05 is α (x axis) and n-2 is the degrees of freedom (df, y axis)

Calculating confidence for a predicted y value

First, calculate the predicted y value at the x value of interest (x_i)

Next, use x_i to obtain S_y

$$y = a + bx_i$$

where *a* is intercept at the y axis and *b* is the slope or rate of change per unit of x axis

$$S_y = \sqrt{MS_{error}\left[\frac{1}{n} + \frac{(x_i - \bar{x})^2}{\Sigma x^2 - \frac{(\Sigma x)^2}{n}}\right]}$$

$$y \pm [(t_{0.05,\, n-2})(s_y)]$$

Table A.2. (two tailed) where 0.05 is α (x axis) and n-2 is the degrees of freedom (df, y axis)

Calculating coefficient of determination (r^2)

$$r^2 = \frac{SS_{regression}}{SS_{total}}$$

this value explains how much of the variability in y is explained by x

0 = no dependence of y on x; 1 = perfect dependence of y on x

TABLE A.2 CRITICAL VALUES OF THE *t* DISTRIBUTION

	α (Two-Tailed)					
df	0.1	0.05	0.02	0.01	0.001	0.0001
1	6.314	12.706	31.821	63.657	636.619	6,366.198
2	2.920	4.303	6.695	9.925	31.598	99.992
3	2.353	3.182	4.541	5.841	12.924	28.000
4	2.132	2.776	3.747	4.604	8.610	15.544
5	2.015	2.571	3.365	4.032	6.869	11.178
6	1.943	2.447	3.143	3.707	5.959	9.082
7	1.895	2.365	2.998	3.499	5.408	7.885
8	1.860	2.306	2.896	3.355	5.041	7.120
9	1.833	2.262	2.821	3.250	4.781	6.594
10	1.812	2.228	2.764	3.169	4.587	6.211
11	1.796	2.201	2.718	3.106	4.437	5.921
12	1.782	2.179	2.681	3.055	4.318	5.694
13	1.771	2.160	2.650	3.012	4.221	5.513
14	1.761	2.145	2.624	2.977	4.140	5.363
15	1.753	2.131	2.602	2.947	4.073	5.239
16	1.746	2.120	2.583	2.921	4.015	5.134
17	1.740	2.110	2.567	2.898	3.965	5.044
18	1.734	2.101	2.552	2.878	3.922	4.966
19	1.729	2.093	2.539	2.861	3.883	4.897
20	1.725	2.086	2.528	2.845	3.850	4.837
21	1.721	2.080	2.518	2.831	3.819	4.784
22	1.717	2.074	2.508	2.819	3.792	4.736
23	1.714	2.069	2.500	2.807	3.767	4.693
24	1.711	2.064	2.492	2.797	3.745	4.654
25	1.708	2.060	2.485	2.787	3.725	4.619
26	1.706	2.056	2.479	2.779	3.707	4.587
27	1.703	2.052	2.473	2.771	3.690	4.558
28	1.701	2.048	2.467	2.763	3.674	4.530
29	1.699	2.045	2.462	2.756	3.659	4.506
30	1.697	2.042	2.457	2.750	3.646	4.482
40	1.684	2.021	2.423	2.704	3.551	4.321
60	1.671	2.000	2.390	2.660	3.460	4.169
100	1.660	1.984	2.364	2.626	3.390	4.053

TABLE A.7 CRITICAL VALUES OF *q* (STUDENTIZED *t*) FOR THE TUKEY TEST (α = 0.05)

Error	Number of Groups (Treatments)								
df	2	3	4	5	6	7	8	9	10
1	17.97	26.98	32.82	37.08	40.41	43.12	45.40	47.36	49.07
2	6.08	8.33	9.80	10.88	11.74	12.44	13.03	13.54	13.99
3	4.50	5.91	6.82	7.50	8.04	8.48	8.85	9.18	9.46
4	3.93	5.04	5.76	6.29	6.71	7.05	7.35	7.60	7.83
5	3.64	4.60	5.22	5.67	6.03	6.33	6.58	6.80	6.99
6	3.46	4.34	4.90	5.30	5.63	5.90	6.12	6.32	6.49
7	3.34	4.16	4.68	5.06	5.36	5.61	5.82	6.00	6.16
8	3.26	4.04	4.53	4.89	5.17	5.40	5.60	5.77	5.92
9	3.20	3.95	4.41	4.76	5.02	5.24	5.43	5.59	5.74
10	3.15	3.88	4.33	4.65	4.91	5.12	5.30	5.46	5.60
11	3.11	3.82	4.26	4.57	4.82	5.03	5.20	5.35	5.49
12	3.08	3.77	4.20	4.51	4.75	4.95	5.12	5.27	5.39
13	3.06	3.73	4.15	4.45	4.69	4.88	5.05	5.19	5.32
14	3.03	3.70	4.11	4.41	4.64	4.83	4.99	5.13	5.25
15	3.01	3.67	4.08	4.37	4.59	4.78	4.94	5.08	5.20
16	3.00	3.65	4.05	4.33	4.56	4.74	4.90	5.03	5.15
17	2.98	3.63	4.02	4.30	4.52	4.70	4.86	4.99	5.11
18	2.97	3.61	4.00	4.28	4.49	4.67	4.82	4.96	5.07
19	2.96	3.59	3.98	4.25	4.47	4.65	4.79	4.92	5.04
20	2.95	3.58	3.96	4.23	4.45	4.62	4.77	4.90	5.01
24	2.92	3.53	3.90	4.17	4.37	4.54	4.68	4.81	4.92
30	2.89	3.49	3.85	4.10	4.30	4.46	4.60	4.72	4.82
40	2.86	3.44	3.79	4.04	4.23	4.39	4.52	4.63	4.73
60	2.83	3.40	3.74	3.98	4.16	4.31	4.44	4.55	4.65
120	2.80	3.36	3.68	3.92	4.10	4.24	4.36	4.47	4.56
∞	2.77	3.31	3.63	3.86	4.03	4.17	4.29	4.39	4.47

TABLE A.6 CRITICAL VALUES OF THE *F* DISTRIBUTION (α = 0.05; df_1 = TREATMENT DEGREES OF FREEDOM, df_2 = ERROR DEGREES OF FREEDOM)

	df_1											
df_2	1	2	3	4	5	6	7	8	9	10	11	12
2	18.5	19.0	19.2	19.3	19.4	19.4	19.4	19.4	19.4	19.4	19.4	19.4
3	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.76	8.74
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.93	5.91
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.71	4.68
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.03	4.00
7	5.59	4.74	4.35	4.12	3.97	3.87	3.77	3.73	3.68	3.64	3.60	3.57
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.31	3.28
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.10	3.07
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.94	2.91
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.82	2.79
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.72	2.69
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.51	2.48
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.31	2.28
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.21	2.16
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.13	2.09
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.04	2.04
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.95	1.92
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91	1.87	1.83