MIE237 Term Test 1

2016-02-09

Examination Type B; Calculator Type 2 Permitted

50 Minutes; 40 Marks Available

Family Name:____

Given Name:____

Student Number:____

This test contains 10 pages. Pages 6–9 are tables. Page 10 is a formula sheet. You can detach the formula sheet if you like, but please don't detach the tables. (Detaching too many pages causes the test to fall apart.) You may use the backs of pages for rough work.

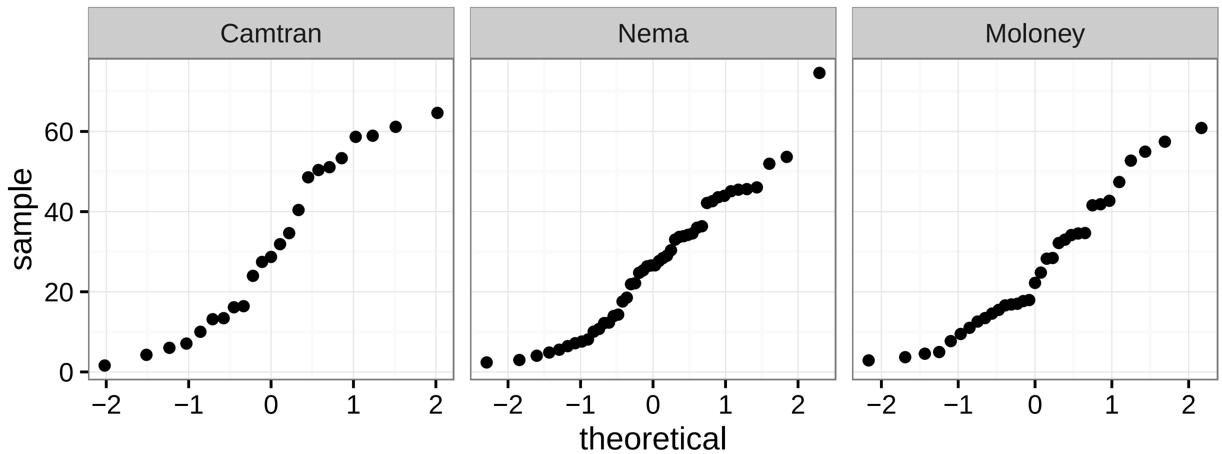
An electricity distribution company (a company that delivers electricity to homes and businesses) has accumulated a dataset related to 102 failed small transformers and wants to analyse some aspects of the data. Here are the first 10 rows of the dataset:

ID	Manufacturer	Size	Age
RY0303	Nema	100KVA	10.7
FD9446	Nema	100KVA	24.7
WZ4786	Moloney	75KVA	11.0
IW5825	Moloney	50KVA	15.5
FZ4835	Moloney	75KVA	41.6
TX9351	Nema	50KVA	36.0
JR0207	Camtran	50KVA	51.1
AB2067	Camtran	100KVA	16.4
BP3860	Moloney	75KVA	28.4
RW5898	Nema	50KVA	21.9

The dataset has 4 variables: `ID, Manufacturer, Size, Age`. The variable `ID` contains the serial number of the transformer. The variable `Manufacturer` contains the manufacturer name, one of: `Camtran, Nema, Moloney`. The variable `Size` contains a description of the transformer's power rating. The variable `Age` contains the age in years of the transformer at the time of its failure.

1. **(15 marks total)** Here is a table of summary statistics with the count, mean age, and standard deviation of age broken down by manufacturer, followed by a normal quantile plot of the ages for each manufacturer.

Manufacturer	Count	Mean Age	SD Age
Camtran	23	31.39	20.91
Nema	46	26.61	16.45
Moloney	33	26.00	16.82



Produce a 95% confidence interval for the difference in mean age at failure between `Camtran` and `Moloney` transformers, commenting on any relevant assumptions you might have needed to make.

2. **(10 marks total)** The company wants to look at the `Manufacturer` and `Size` variables. Here is a summary table with counts by these two variables, followed by `R` output for the χ^2 test of independence with some values removed (replaced with `MISSING`).

	100KVA	75KVA	50KVA	Sum
Camtran	9	3	11	23
Nema	9	17	20	46
Moloney	8	18	7	33
Sum	26	38	38	102

```
##
##  Pearson's Chi-squared test
##
## data:  tx$Size and tx$Manufacturer
## X-squared = 12.049, df = MISSING, p-value = MISSING
```

a. **(3 marks)** Produce a 95% confidence interval for the proportion of transformers that are manufactured by Nema, commenting on any relevant assumptions you might have needed to make.

b. **(2 marks)** Compute the "expected cell count" for the top left cell (corresponding to `Camtran` and `100KVA`).

c. **(2 marks)** How many out of the 9 expected cell counts would you need to calculate using multiplication and division of marginal totals before you can simply use addition and subtraction to produce the rest?

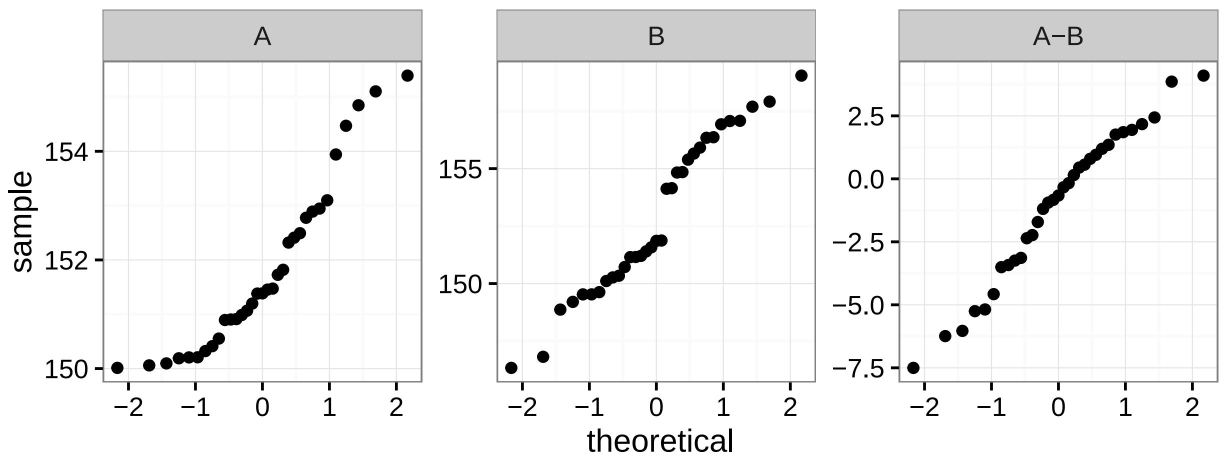
d. **(3 marks)** Perform the test of independence with null hypothesis (informally) expressed as: H_0 : `Manufacturer` and `Size` are independent, commenting on any relevant assumptions you might have needed to make.

3. **(10 marks total)** The company happens to still have all the `Moloney` transformers in storage and decides to do some electrical testing on two of the "windings" (essentially, a wire wound around a metal core—the details don't matter) in each of these transformers. Let's call the windings `A` and `B` within each unit. A current is passed through each winding and the amount of heat generated is measured. (If you are a transformer expert and this makes no sense, this is all made up, and please forgive me.)

A summer student working at the company produces the following summaries of the data gathered, consisting of: mean and standard deviation for each of the `A` and `B` winding experiments, and the standard deviation of the unit-by-unit differences between `A` and `B` experiments.

Count	A Temp Mean	A Temp SD	B Temp Mean	B Temp SD	A–B Diff Temp SD
33	151.82	1.56	152.88	1.56	3.01

Here are the normal quantile plots for the `A` and `B` winding experiments and also for the `A-B` differences.



Perform the appropriate hypothesis test to evaluate if there is a difference in temperature between `A` and `B` winding experiments.

4. **(5 marks total)** Consider the simple linear regression model $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ with $\varepsilon_i \sim N(0, \sigma^2)$. The least squares estimators for β_0 and β_1 are on the aid sheet—you'll need them here. (In this question for economy of notation I've used lowercase y to refer to "data" and "random variable" interchangeably.)

a. **(1 mark)** Show that the fitted regression line $y = \hat{\beta}_0 + \hat{\beta}_1 x$ always passes through the point $(\overline{x}, \overline{y})$ for any dataset $\{(y_1, x_1), \ldots, (y_n, x_n)\}$.

b. **(2 marks)** Show that $E(\overline{y}) = \beta_0 + \beta_1 \overline{x}$.

c. **(2 marks)** Show that $E(\hat{\beta}_1) = \beta_1$.

Standard Normal Probabilities $P(Z \leq z)$

	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
-3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
-3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
-3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
-3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
-2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
-1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
-1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
-1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
-0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
-0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
-0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
-0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641

Standard Normal Probabilities $P(Z \le z)$

	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

Upper tail probabilities for t_ν distributions $P(t_\nu \geq t)$

df	0.3	0.2	0.15	0.1	0.05	0.025	0.02	0.015	0.01	0.0075	0.005	0.0025	0.0005
11	0.540	0.876	1.088	1.363	1.796	2.201	2.328	2.491	2.718	2.879	3.106	3.497	4.437
12	0.539	0.873	1.083	1.356	1.782	2.179	2.303	2.461	2.681	2.836	3.055	3.428	4.318
13	0.538	0.870	1.079	1.350	1.771	2.160	2.282	2.436	2.650	2.801	3.012	3.372	4.221
14	0.537	0.868	1.076	1.345	1.761	2.145	2.264	2.415	2.624	2.771	2.977	3.326	4.140
15	0.536	0.866	1.074	1.341	1.753	2.131	2.249	2.397	2.602	2.746	2.947	3.286	4.073
16	0.535	0.865	1.071	1.337	1.746	2.120	2.235	2.382	2.583	2.724	2.921	3.252	4.015
17	0.534	0.863	1.069	1.333	1.740	2.110	2.224	2.368	2.567	2.706	2.898	3.222	3.965
18	0.534	0.862	1.067	1.330	1.734	2.101	2.214	2.356	2.552	2.689	2.878	3.197	3.922
19	0.533	0.861	1.066	1.328	1.729	2.093	2.205	2.346	2.539	2.674	2.861	3.174	3.883
20	0.533	0.860	1.064	1.325	1.725	2.086	2.197	2.336	2.528	2.661	2.845	3.153	3.850
21	0.532	0.859	1.063	1.323	1.721	2.080	2.189	2.328	2.518	2.649	2.831	3.135	3.819
22	0.532	0.858	1.061	1.321	1.717	2.074	2.183	2.320	2.508	2.639	2.819	3.119	3.792
23	0.532	0.858	1.060	1.319	1.714	2.069	2.177	2.313	2.500	2.629	2.807	3.104	3.768
24	0.531	0.857	1.059	1.318	1.711	2.064	2.172	2.307	2.492	2.620	2.797	3.091	3.745
25	0.531	0.856	1.058	1.316	1.708	2.060	2.167	2.301	2.485	2.612	2.787	3.078	3.725
26	0.531	0.856	1.058	1.315	1.706	2.056	2.162	2.296	2.479	2.605	2.779	3.067	3.707
27	0.531	0.855	1.057	1.314	1.703	2.052	2.158	2.291	2.473	2.598	2.771	3.057	3.690
28	0.530	0.855	1.056	1.313	1.701	2.048	2.154	2.286	2.467	2.592	2.763	3.047	3.674
29	0.530	0.854	1.055	1.311	1.699	2.045	2.150	2.282	2.462	2.586	2.756	3.038	3.659
30	0.530	0.854	1.055	1.310	1.697	2.042	2.147	2.278	2.457	2.581	2.750	3.030	3.646
31	0.530	0.853	1.054	1.309	1.696	2.040	2.144	2.275	2.453	2.576	2.744	3.022	3.633
32	0.530	0.853	1.054	1.309	1.694	2.037	2.141	2.271	2.449	2.571	2.738	3.015	3.622
33	0.530	0.853	1.053	1.308	1.692	2.035	2.138	2.268	2.445	2.566	2.733	3.008	3.611
34	0.529	0.852	1.052	1.307	1.691	2.032	2.136	2.265	2.441	2.562	2.728	3.002	3.601
35	0.529	0.852	1.052	1.306	1.690	2.030	2.133	2.262	2.438	2.558	2.724	2.996	3.591
36	0.529	0.852	1.052	1.306	1.688	2.028	2.131	2.260	2.434	2.555	2.719	2.990	3.582
37	0.529	0.851	1.051	1.305	1.687	2.026	2.129	2.257	2.431	2.551	2.715	2.985	3.574
38	0.529	0.851	1.051	1.304	1.686	2.024	2.127	2.255	2.429	2.548	2.712	2.980	3.566
39	0.529	0.851	1.050	1.304	1.685	2.023	2.125	2.252	2.426	2.545	2.708	2.976	3.558
40	0.529	0.851	1.050	1.303	1.684	2.021	2.123	2.250	2.423	2.542	2.704	2.971	3.551
41	0.529	0.850	1.050	1.303	1.683	2.020	2.121	2.248	2.421	2.539	2.701	2.967	3.544
42	0.528	0.850	1.049	1.302	1.682	2.018	2.120	2.246	2.418	2.537	2.698	2.963	3.538
43	0.528	0.850	1.049	1.302	1.681	2.017	2.118	2.244	2.416	2.534	2.695	2.959	3.532
44	0.528	0.850	1.049	1.301	1.680	2.015	2.116	2.243	2.414	2.532	2.692	2.956	3.526
45	0.528	0.850	1.049	1.301	1.679	2.014	2.115	2.241	2.412	2.529	2.690	2.952	3.520
46	0.528	0.850	1.048	1.300	1.679	2.013	2.114	2.239	2.410	2.527	2.687	2.949	3.515
47	0.528	0.849	1.048	1.300	1.678	2.012	2.112	2.238	2.408	2.525	2.685	2.946	3.510
48	0.528	0.849	1.048	1.299	1.677	2.011	2.111	2.237	2.407	2.523	2.682	2.943	3.505
49	0.528	0.849	1.048	1.299	1.677	2.010	2.110	2.235	2.405	2.521	2.680	2.940	3.500
50	0.528	0.849	1.047	1.299	1.676	2.009	2.109	2.234	2.403	2.519	2.678	2.937	3.496
51	0.528	0.849	1.047	1.298	1.675	2.008	2.108	2.233	2.402	2.518	2.676	2.934	3.492
52	0.528	0.849	1.047	1.298	1.675	2.007	2.107	2.231	2.400	2.516	2.674	2.932	3.488
53	0.528	0.848	1.047	1.298	1.674	2.006	2.106	2.230	2.399	2.514	2.672	2.929	3.484
54	0.528	0.848	1.046	1.297	1.674	2.005	2.105	2.229	2.397	2.513	2.670	2.927	3.480
55	0.527	0.848	1.046	1.297	1.673	2.004	2.104	2.228	2.396	2.511	2.668	2.925	3.476
56	0.527	0.848	1.046	1.297	1.673	2.003	2.103	2.227	2.395	2.510	2.667	2.923	3.473
57	0.527	0.848	1.046	1.297	1.672	2.002	2.102	2.226	2.394	2.508	2.665	2.920	3.470
58	0.527	0.848	1.046	1.296	1.672	2.002	2.101	2.225	2.392	2.507	2.663	2.918	3.466
59	0.527	0.848	1.046	1.296	1.671	2.001	2.100	2.224	2.391	2.506	2.662	2.916	3.463
60	0.527	0.848	1.045	1.296	1.671	2.000	2.099	2.223	2.390	2.504	2.660	2.915	3.460
120	0.526	0.845	1.041	1.289	1.658	1.980	2.076	2.196	2.358	2.468	2.617	2.860	3.373
∞	0.524	0.842	1.036	1.282	1.645	1.960	2.054	2.170	2.326	2.432	2.576	2.807	3.291

Upper tail probabilities for χ^2_ν distributions $P(\chi^2_\nu \geq \chi^2)$

df	0.3	0.2	0.15	0.1	0.05	0.025	0.02	0.015	0.01	0.0075	0.005	0.0025	0.0005
1	1.074	1.642	2.072	2.706	3.841	5.024	5.412	5.916	6.635	7.149	7.879	9.141	12.116
2	2.408	3.219	3.794	4.605	5.991	7.378	7.824	8.399	9.210	9.786	10.597	11.983	15.202
3	3.665	4.642	5.317	6.251	7.815	9.348	9.837	10.465	11.345	11.966	12.838	14.320	17.730
4	4.878	5.989	6.745	7.779	9.488	11.143	11.668	12.339	13.277	13.937	14.860	16.424	19.997
5	6.064	7.289	8.115	9.236	11.070	12.833	13.388	14.098	15.086	15.780	16.750	18.386	22.105
6	7.231	8.558	9.446	10.645	12.592	14.449	15.033	15.777	16.812	17.537	18.548	20.249	24.103
7	8.383	9.803	10.748	12.017	14.067	16.013	16.622	17.398	18.475	19.229	20.278	22.040	26.018
8	9.524	11.030	12.027	13.362	15.507	17.535	18.168	18.974	20.090	20.870	21.955	23.774	27.868
9	10.656	12.242	13.288	14.684	16.919	19.023	19.679	20.513	21.666	22.471	23.589	25.462	29.666
10	11.781	13.442	14.534	15.987	18.307	20.483	21.161	22.021	23.209	24.038	25.188	27.112	31.420
11	12.899	14.631	15.767	17.275	19.675	21.920	22.618	23.503	24.725	25.576	26.757	28.729	33.137
12	14.011	15.812	16.989	18.549	21.026	23.337	24.054	24.963	26.217	27.090	28.300	30.318	34.821
13	15.119	16.985	18.202	19.812	22.362	24.736	25.472	26.403	27.688	28.582	29.819	31.883	36.478
14	16.222	18.151	19.406	21.064	23.685	26.119	26.873	27.827	29.141	30.055	31.319	33.426	38.109
15	17.322	19.311	20.603	22.307	24.996	27.488	28.259	29.235	30.578	31.511	32.801	34.950	39.719
16	18.418	20.465	21.793	23.542	26.296	28.845	29.633	30.629	32.000	32.952	34.267	36.456	41.308
17	19.511	21.615	22.977	24.769	27.587	30.191	30.995	32.011	33.409	34.378	35.718	37.946	42.879
18	20.601	22.760	24.155	25.989	28.869	31.526	32.346	33.382	34.805	35.793	37.156	39.422	44.434
19	21.689	23.900	25.329	27.204	30.144	32.852	33.687	34.742	36.191	37.195	38.582	40.885	45.973
20	22.775	25.038	26.498	28.412	31.410	34.170	35.020	36.093	37.566	38.588	39.997	42.336	47.498
21	23.858	26.171	27.662	29.615	32.671	35.479	36.343	37.434	38.932	39.970	41.401	43.775	49.011
22	24.939	27.301	28.822	30.813	33.924	36.781	37.659	38.768	40.289	41.343	42.796	45.204	50.511
23	26.018	28.429	29.979	32.007	35.172	38.076	38.968	40.094	41.638	42.707	44.181	46.623	52.000
24	27.096	29.553	31.132	33.196	36.415	39.364	40.270	41.413	42.980	44.064	45.559	48.034	53.479
25	28.172	30.675	32.282	34.382	37.652	40.646	41.566	42.725	44.314	45.413	46.928	49.435	54.947
26	29.246	31.795	33.429	35.563	38.885	41.923	42.856	44.031	45.642	46.756	48.290	50.829	56.407
27	30.319	32.912	34.574	36.741	40.113	43.195	44.140	45.331	46.963	48.091	49.645	52.215	57.858
28	31.391	34.027	35.715	37.916	41.337	44.461	45.419	46.626	48.278	49.421	50.993	53.594	59.300
29	32.461	35.139	36.854	39.087	42.557	45.722	46.693	47.915	49.588	50.744	52.336	54.967	60.735
30	33.530	36.250	37.990	40.256	43.773	46.979	47.962	49.199	50.892	52.062	53.672	56.332	62.162

Two Samples

Model: $Y_{ij} = \mu_i + \varepsilon_{ij}$ with $i \in \{1, 2\}$ and ε_{ij} i.i.d. $N(0, \sigma^2)$.

Pooled sample variance:

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

Test Statistic:

$$T = \frac{(\overline{Y}_{1\cdot} - \overline{Y}_{2\cdot}) - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2}$$

$(1 - \alpha) \cdot 100\%$ C.I. is

$$\overline{Y}_{1\cdot} - \overline{Y}_{2\cdot} \pm t_{n_1+n_2-2,\alpha/2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Checking assumptions: normal plots and 3:1 SD ratio for equal variance assumption.

If observations are really "paired", use one-sample procedures on the paired differences using this fact:

$$\frac{\overline{Y}_d - \mu_d}{S_d/\sqrt{n}} \sim t_{n-1}$$

where $\overline{Y}_d$ is the sample average of the differences, μ_d is the mean difference between the two populations, S_d is the sample standard deviation of the differences and n is the number of paired observations.

Simple Linear Regression

Model: $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ with $\varepsilon_i \sim N(0, \sigma^2)$

Analysis: $\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}$

$\hat{\beta}_1 = S_{xy}/S_{xx}$

$S_{xy} = \sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y})$

$S_{xx} = \sum_{i=1}^n (x_i - \overline{x})^2$

$\text{Var}\left(\hat{\beta}_1\right) = \frac{\sigma^2}{S_{xx}}$

Fitted value at x_i is $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$

SS decomposition details:

$$\begin{array}{ccccc}
\sum_{i=1}^n \left(Y_i - \overline{Y}\right)^2 & = & \sum_{i=1}^n \left(\hat{Y}_i - \overline{Y}\right)^2 & + & \sum_{i=1}^n \left(Y_i - \hat{Y}_i\right)^2 \\
\text{SST} & = & \text{SSR} & + & \text{SSE} \\
n-1 \text{ d.f.} & = & 1 \text{ d.f.} & + & n-2 \text{ d.f.}
\end{array}$$

Test statistic for β_1:

$$T = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{MSE/S_{xx}}} \sim t_{n-2}$$

The denominator is called the "standard error" of $\hat{\beta}_1$

$(1 - \alpha) \cdot 100\%$ C.I. for β_1 is:

$$\hat{\beta}_1 \pm t_{n-2,\alpha/2} \sqrt{\frac{MSE}{S_{xx}}}$$

Alternate approach for $H_0 : \beta_1 = 0$ versus $H_1 : \beta_1 \neq 0$ uses (again... $T^2 = F$):

$$F = \frac{SSR/1}{SSE/(n-2)} = \frac{MSR}{MSE} \sim F_{1,n-2}$$

CI for Proportion

$$\hat{p} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}$$

Approximately valid if $n\hat{p}$ and $n(1 - \hat{p})$ exceed 5.

Test for Independence

Given an observed r by c contingency table compute the expected cell counts E_i by keeping the row and column totals fixed. The test statistic is then:

$$\chi^2_{obs} = \sum_{i=1} \frac{(O_i - E_i)^2}{E_i}$$

where the sum is over all entries in the table, and has an approximate χ^2 distribution with $(r - 1) \times (c - 1)$ degrees of freedom as long as the E_i all exceed 5.