Basic Statistics Formulas

Population Measures

$$\text{Mean } \mu = \frac{1}{n}\sum x_i \tag{1}$$

$$\text{Variance } \sigma^2 = \frac{1}{n}\sum (x_i - \overline{x})^2 \tag{2}$$

$$\text{Standard Deviation } \sigma = \sqrt{\frac{1}{n}\sum (x_i - \overline{x})^2} \tag{3}$$

Sampling

$$\text{Sample mean } \overline{x} = \frac{1}{n}\sum x_i \tag{4}$$

$$\text{Sample variance } s_x^2 = \frac{1}{n-1}\sum (x_i - \overline{x})^2 \tag{5}$$

$$\text{Std. Deviation } s_x = \sqrt{\frac{1}{n-1}\sum (x_i - \overline{x})^2} \tag{6}$$

$$\text{z-score } z = \frac{x-\mu}{\sigma} \tag{7}$$

Correlation $r =$

$$\frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{(x_i - \overline{x})}{s_x}\right)\left(\frac{(y_i - \overline{y})}{s_y}\right) \tag{8}$$

Linear Regression

$$\text{Line } \hat{y} = a + bx \tag{9}$$

$$b = r\frac{s_y}{s_x}, a = \overline{y} - b\overline{x} \tag{10}$$

$$s = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n}(y_i - \hat{y})^2} \tag{11}$$

$$SE_b = \frac{s}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}} \tag{12}$$

$$\text{To test } H_0 : b = 0, \text{ use } t = \frac{b}{SE_b} \tag{13}$$

$$CI = b \pm t^* SE_b \tag{14}$$

Probability

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B) \tag{15}$$

$$P(\text{not } A) = 1 - P(A) \tag{16}$$

$$P(A \text{ and } B) = P(A)P(B) \text{ (independent)} \tag{17}$$

$$P(B|A) = P(A \text{ and } B)/P(A) \tag{18}$$

$$0! = 1; n! = 1 \times 2 \times 3 \cdots \times (n-1) \times n \tag{19}$$

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \tag{20}$$

Binomial Distribution :

$$P(\mathcal{X} = k) = \binom{n}{k} p^k (1-p)^{n-k} \tag{21}$$

$$\mu = np, \ \sigma = \sqrt{np(1-p)} \tag{22}$$

One-Sample z-statistic

$$\text{To test } H_0 : \mu = \mu_0 \text{ use} z = \frac{\overline{z} - \mu_0}{\sigma/\sqrt{n}} \tag{23}$$

$$\text{Confidence Interval for } \mu = \overline{x} \pm z^* \frac{\sigma}{\sqrt{n}} \tag{24}$$

$$\text{Margin of Error } ME = z^* \frac{\sigma}{\sqrt{n}} \tag{25}$$

$$\text{Minimum sample size } n \geq \left[\frac{z^*\sigma}{ME}\right]^2 \tag{26}$$

One-Sample t-statistic

$$SEM = \frac{s_x}{\sqrt{n}}, \ t = \frac{\overline{x} - \mu}{s_x/\sqrt{n}} \tag{27}$$

$$\text{Confidence Interval} = \overline{x} \pm t^* \frac{s_x}{\sqrt{n}} \tag{28}$$

Two-Sample t-statistic

$$t = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \tag{29}$$

$$\text{Conf. Interval} = (\overline{x}_1 - \overline{x}_2) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \tag{30}$$

Sample Proportions

$$\mu_{\hat{p}} = p, \ \sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} \tag{31}$$

$$\text{Conf. Int.} = \hat{p} \pm z^*(SE) \tag{32}$$

$$\text{SE} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \tag{33}$$

$$\text{sample size } n > \left[\frac{z^*}{ME}\right]^2 p^*(1-p^*) \tag{34}$$

$$\text{To test } H_0 : p = p_0, \text{ use } z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \tag{35}$$

Two-Sample Proportions

$$SE = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \tag{36}$$

$$CI = (\hat{p}_1 - \hat{p}_2) \pm z^*(SE) \tag{37}$$

$$\text{To test } H_0 : p_1 = p_2, \text{ use} \tag{38}$$

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \tag{39}$$

$$\hat{p} = \frac{X_1 + X_2}{n_1 + n_2}, \ X_i = \text{ successes} \tag{40}$$

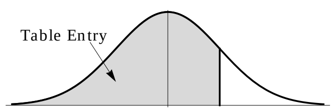
Chi-Square Statistic

$$\chi^2 = \sum_{i=1}^{n}\frac{(o_i - e_i)^2}{e_i} \tag{41}$$

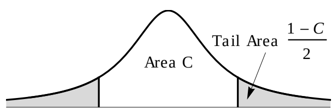
$o_i = \text{observed}, e_i = \text{expected}$

Central Limit Theorem

$$s_{\overline{x}} \to \frac{\sigma}{\sqrt{n}} \text{ as } n \to \infty \tag{42}$$



Standard Normal Cumulative Proportions (below)



t-Distribution Critical Values (to right)

Standard Normal Cumulative Proportions

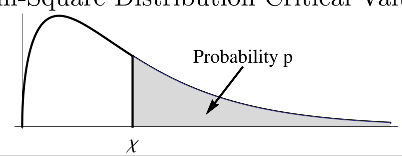
	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
-3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
-3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
-3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
-3	0.0013	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	0.0228	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	0.1587	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	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641

	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
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	0.8413	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	0.9772	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	0.9987	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

t-Distribution Cumulative Proportions

	Confidence Level C									
df	50%	60%	70%	80%	90%	95%	96%	98%	99%	99.8%
1	1	1.376	1.963	3.078	6.314	12.706	15.895	31.821	63.657	318.309
2	0.816	1.061	1.386	1.886	2.92	4.303	4.849	6.965	9.925	22.327
3	0.765	0.978	1.25	1.638	2.353	3.182	3.482	4.541	5.841	10.215
4	0.741	0.941	1.19	1.533	2.132	2.776	2.999	3.747	4.604	7.173
5	0.727	0.92	1.156	1.476	2.015	2.571	2.757	3.365	4.032	5.893
6	0.718	0.906	1.134	1.44	1.943	2.447	2.612	3.143	3.707	5.208
7	0.711	0.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.785
8	0.706	0.889	1.108	1.397	1.86	2.306	2.449	2.896	3.355	4.501
9	0.703	0.883	1.1	1.383	1.833	2.262	2.398	2.821	3.25	4.297
10	0.7	0.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	4.144
11	0.697	0.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	4.025
12	0.695	0.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.93
13	0.694	0.87	1.079	1.35	1.771	2.16	2.282	2.65	3.012	3.852
14	0.692	0.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.787
15	0.691	0.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.733
16	0.69	0.865	1.071	1.337	1.746	2.12	2.235	2.583	2.921	3.686
17	0.689	0.863	1.069	1.333	1.74	2.11	2.224	2.567	2.898	3.646
18	0.688	0.862	1.067	1.33	1.734	2.101	2.214	2.552	2.878	3.61
19	0.688	0.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861	3.579
20	0.687	0.86	1.064	1.325	1.725	2.086	2.197	2.528	2.845	3.552
21	0.686	0.859	1.063	1.323	1.721	2.08	2.189	2.518	2.831	3.527
22	0.686	0.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819	3.505
23	0.685	0.858	1.06	1.319	1.714	2.069	2.177	2.5	2.807	3.485
24	0.685	0.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797	3.467
25	0.684	0.856	1.058	1.316	1.708	2.06	2.167	2.485	2.787	3.45
30	0.683	0.854	1.055	1.31	1.697	2.042	2.147	2.457	2.75	3.385
40	0.681	0.851	1.05	1.303	1.684	2.021	2.123	2.423	2.704	3.307
50	0.679	0.849	1.047	1.299	1.676	2.009	2.109	2.403	2.678	3.261
60	0.679	0.848	1.045	1.296	1.671	2	2.099	2.39	2.66	3.232
80	0.678	0.846	1.043	1.292	1.664	1.99	2.088	2.374	2.639	3.195
100	0.677	0.845	1.042	1.29	1.66	1.984	2.081	2.364	2.626	3.174
1000	0.675	0.842	1.037	1.282	1.646	1.962	2.056	2.33	2.581	3.098
z^*	0.674	0.842	1.036	1.282	1.645	1.960	2.054	2.326	2.576	3.090
1-Sided P	0.25	0.2	0.15	0.1	0.05	0.025	0.02	0.01	0.005	0.001
2-Sided P	0.5	0.4	0.3	0.2	0.1	0.05	0.04	0.02	0.01	0.002

Chi-Square Distribution Critical Values



	p									
df	0.25	0.20	0.10	0.05	0.025	0.02	0.01	0.005	0.0025	0.001
1	1.32	1.64	2.71	3.84	5.02	5.41	6.63	7.88	9.14	10.83
2	2.77	3.22	4.61	5.99	7.38	7.82	9.21	10.60	11.98	13.82
3	4.11	4.64	6.25	7.81	9.35	9.84	11.34	12.84	14.32	16.27
4	5.39	5.99	7.78	9.49	11.14	11.67	13.28	14.86	16.42	18.47
5	6.63	7.29	9.24	11.07	12.83	13.39	15.09	16.75	18.39	20.52
6	7.84	8.56	10.64	12.59	14.45	15.03	16.81	18.55	20.25	22.46
7	9.04	9.80	12.02	14.07	16.01	16.62	18.48	20.28	22.04	24.32
8	10.22	11.03	13.36	15.51	17.53	18.17	20.09	21.95	23.77	26.12
9	11.39	12.24	14.68	16.92	19.02	19.68	21.67	23.59	25.46	27.88
10	12.55	13.44	15.99	18.31	20.48	21.16	23.21	25.19	27.11	29.59
11	13.70	14.63	17.28	19.68	21.92	22.62	24.72	26.76	28.73	31.26
12	14.85	15.81	18.55	21.03	23.34	24.05	26.22	28.30	30.32	32.91
13	15.98	16.98	19.81	22.36	24.74	25.47	27.69	29.82	31.88	34.53
14	17.12	18.15	21.06	23.68	26.12	26.87	29.14	31.32	33.43	36.12
15	18.25	19.31	22.31	25.00	27.49	28.26	30.58	32.80	34.95	37.70
16	19.37	20.47	23.54	26.30	28.85	29.63	32.00	34.27	36.46	39.25
17	20.49	21.61	24.77	27.59	30.19	31.00	33.41	35.72	37.95	40.79
18	21.60	22.76	25.99	28.87	31.53	32.35	34.81	37.16	39.42	42.31
19	22.72	23.90	27.20	30.14	32.85	33.69	36.19	38.58	40.88	43.82
20	23.83	25.04	28.41	31.41	34.17	35.02	37.57	40.00	42.34	45.31
21	24.93	26.17	29.62	32.67	35.48	36.34	38.93	41.40	43.78	46.80
22	26.04	27.30	30.81	33.92	36.78	37.66	40.29	42.80	45.20	48.27
23	27.14	28.43	32.01	35.17	38.08	38.97	41.64	44.18	46.62	49.73
24	28.24	29.55	33.20	36.42	39.36	40.27	42.98	45.56	48.03	51.18
25	29.34	30.68	34.38	37.65	40.65	41.57	44.31	46.93	49.44	52.62
30	34.80	36.25	40.26	43.77	46.98	47.96	50.89	53.67	56.33	59.70
40	45.62	47.27	51.81	55.76	59.34	60.44	63.69	66.77	69.70	73.40
50	56.33	58.16	63.17	67.50	71.42	72.61	76.15	79.49	82.66	86.66
60	66.98	68.97	74.40	79.08	83.30	84.58	88.38	91.95	95.34	99.61
80	88.13	90.41	96.58	101.88	106.63	108.07	112.33	116.32	120.10	124.84
100	109.14	111.67	118.50	124.34	129.56	131.14	135.81	140.17	144.29	149.45

Frequently Used Statistics Formulas and Tables

Chapter 2

$$\text{Class Width} = \frac{\text{highest value - lowest value}}{\text{number classes}} \quad \text{(increase to next integer)}$$

$$\text{Class Midpoint} = \frac{\text{upper limit + lower limit}}{2}$$

Chapter 3

$n =$ sample size

$N =$ population size

$f =$ frequency

$\Sigma = sum$

$w = weight$

Sample mean: $\bar{x} = \dfrac{\sum x}{n}$

Population mean: $\mu = \dfrac{\sum x}{N}$

Weighted mean: $\bar{x} = \dfrac{\sum(w \bullet x)}{\sum w}$

Mean for frequency table: $\bar{x} = \dfrac{\sum(f \bullet x)}{\sum f}$

$$\text{Midrange} = \frac{\text{highest value + lowest value}}{2}$$

Range = Highest value - Lowest value

Sample standard deviation: $s = \sqrt{\dfrac{\sum(x-\bar{x})^2}{n-1}}$

Population standard deviation: $\sigma = \sqrt{\dfrac{\sum(x-\mu)^2}{N}}$

Sample variance: s^2

Population variance: σ^2

Chapter 3

Limits for Unusual Data

Below : μ - 2σ

Above: $\mu + 2\sigma$

Empirical Rule

About 68%: μ-σ to $\mu + \sigma$

About 95%: μ-2σ to $\mu + 2\sigma$

About 99.7%: μ-3σ to $\mu + 3\sigma$

Sample coefficient of variation: $CV = \dfrac{s}{\bar{x}} \bullet 100\%$

Population coefficient of variation: $CV = \dfrac{\sigma}{\mu} \bullet 100\%$

Sample standard deviation for frequency table:

$$s = \sqrt{\frac{n\,[\ \sum(f \bullet x^2)\ ] - [\ \sum(f \bullet x)\ ]^2}{n\,(n-1)}}$$

Sample z-score: $z = \dfrac{x-\bar{x}}{s}$

Population z-score: $z = \dfrac{x-\mu}{\sigma}$

Interquartile Range: (IQR) $= Q_3 - Q_1$

Modified Box Plot Outliers

lower limit: Q_1 - 1.5 (IQR)

upper limit: Q_3 + 1.5 (IQR)

Chapter 4

Probability of the complement of event A
$P(not\ A) = 1 - P(A)$

Multiplication rule for independent events
$P(A\ and\ B) = P(A) \bullet P(B)$

General multiplication rules
$P(A\ and\ B) = P(A) \bullet P(B,\ given\ A)$
$P(A\ and\ B) = P(A) \bullet P(A,\ given\ B)$

Addition rule for mutually exclusive events
$P(A\ or\ B) = P(A) + P(B)$

General addition rule
$P(A\ or\ B) = P(A) + P(B) - P(A\ and\ B)$

Permutation rule: $_nP_r = \dfrac{n!}{(n-r)!}$

Combination rule: $_nC_r = \dfrac{n!}{r!(n-r)!}$

<u>Permutation and Combination on TI 83/84</u>

n [Math] [PRB] [nPr] [enter] r

n [Math] [PRB] [nCr] [enter] r

Note: textbooks and formula sheets interchange "r" and "x" for number of successes

Chapter 5

Discrete Probability Distributions:

Mean of a discrete probability distribution:
$\mu = \sum[x \bullet P(x)]$

Standard deviation of a probability distribution:

$\sigma = \sqrt{\sum[x^2 \bullet P(x)] - \mu^2}$

Binomial Distributions

r = number of successes (or x)

p = probability of success

q = probability of failure

$q = 1 - p$ $\quad p + q = 1$

Binomial probability distribution
$P(r) = {}_nC_r\, p^r q^{n-r}$

Mean: $\mu = np$

Standard deviation: $\sigma = \sqrt{npq}$

Poisson Distributions

r = number of successes (or x)

μ = mean number of successes
(over a given interval)

Poisson probability distribution

$$P(r) = \frac{e^{-\mu}\mu^r}{r!}$$

$e \approx 2.71828$

$\mu = mean$ (over some interval)

$\sigma = \sqrt{\mu}$

$\sigma^2 = \mu$

Chapter 6

Normal Distributions

Raw score: $x = z\sigma + \mu$

Standard score: $z = \frac{x-\mu}{\sigma}$

Mean of $\overline{x}$ distribution: $\mu_{\overline{x}} = \mu$

Standard deviation of $\overline{x}$ distribtuion: $\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}}$

(standard error)

Standard score for $\overline{x}$: $z = \frac{\overline{x}-\mu}{\sigma/\sqrt{n}}$

Chapter 7

One Sample Confidence Interval

for proportions (p): $(np > 5 \text{ and } nq > 5)$

$\hat{p} - E < p < \hat{p} + E$

where $E = z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}$

$\hat{p} = \frac{r}{n}$

for means (μ) when σ is known:

$\overline{x} - E < \mu < \overline{x} + E$

where $E = z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$

for means (μ) when σ is unknown:

$\overline{x} - E < \mu < \overline{x} + E$

where $E = t_{\alpha/2}\frac{s}{\sqrt{n}}$

with $d.f. = n-1$

for variance (σ^2): $\frac{(n-1)s^2}{\chi_R^2} < \sigma^2 < \frac{(n-1)s^2}{\chi_L^2}$

with $d.f. = n-1$

Chapter 7

Confidence Interval: Point estimate ± error

Point estimate = $\dfrac{\text{Upper limit + Lower limit}}{2}$

Error = $\dfrac{\text{Upper limit - Lower limit}}{2}$

Sample Size for Estimating

means:

$n = \left(\frac{z_{\alpha/2}\sigma}{E}\right)^2$

proportions:

$n = \hat{p}\hat{q}\left(\frac{z_{\alpha/2}}{E}\right)^2$ with preliminary estimate for p

$n = 0.25\left(\frac{z_{\alpha/2}}{E}\right)^2$ without preliminary estimate for p

variance or standard deviation:
*see table 7-2 (last page of formula sheet)

Confidence Intervals

Level of Confidence	z-value $(z_{\alpha/2})$
70%	1.04
75%	1.15
80%	1.28
85%	1.44
90%	1.645
95%	1.96
98%	2.33
99%	2.58

Chapter 8

One Sample Hypothesis Testing

for p ($np > 5$ and $nq > 5$): $\quad z = \dfrac{\hat{p} - p}{\sqrt{pq/n}}$

where $q = 1 - p$; $\hat{p} = r/n$

for μ (σ known): $\quad z = \dfrac{\overline{x} - \mu}{\sigma/\sqrt{n}}$

for μ (σ unknown): $\quad t = \dfrac{\overline{x} - \mu}{s/\sqrt{n}}$ with $d.f. = n - 1$

for σ^2: $\quad \chi^2 = \dfrac{(n-1)s^2}{\sigma^2}$ with $d.f. = n - 1$

Chapter 9

Two Sample Confidence Intervals and Tests of Hypotheses

Difference of Proportions ($p_1 - p_2$)

Confidence Interval:

$$(\hat{p}_1 - \hat{p}_2) - E < (p_1 - p_2) < (\hat{p}_1 - \hat{p}_2) + E$$

where $E = z_{\alpha/2}\sqrt{\dfrac{\hat{p}_1\hat{q}_1}{n_1} + \dfrac{\hat{p}_2\hat{q}_2}{n_2}}$

$\hat{p}_1 = r_1/n_1$; $\hat{p}_2 = r_2/n_2$ and $\hat{q}_1 = 1 - \hat{p}_1$; $\hat{q}_2 = 1 - \hat{p}_2$

Hypothesis Test:

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\dfrac{\overline{p}\,\overline{q}}{n_1} + \dfrac{\overline{p}\,\overline{q}}{n_2}}}$$

where the pooled proportion is $\overline{p}$

$$\overline{p} = \frac{r_1 + r_2}{n_1 + n_2} \text{ and } \overline{q} = 1 - \overline{p}$$

$\hat{p}_1 = r_1/n_1$; $\hat{p}_2 = r_2/n_2$

Chapter 9

Difference of means $\mu_1 - \mu_2$ (independent samples)

Confidence Interval when σ_1 and σ_2 are known

$$(\overline{x}_1 - \overline{x}_2) - E < (\mu_1 - \mu_2) < (\overline{x}_1 - \overline{x}_2) + E$$

where $E = z_{\alpha/2}\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}$

Hypothesis Test when σ_1 and σ_2 are known

$$z = \frac{(\overline{x}_1 - \overline{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$$

Confidence Interval when σ_1 and σ_2 are unknown

$$(\overline{x}_1 - \overline{x}_2) - E < (\mu_1 - \mu_2) < (\overline{x}_1 - \overline{x}_2) + E$$

$E = t_{\alpha/2}\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}$

with $d.f.$ = smaller of $n_1 - 1$ and $n_2 - 1$

Hypothesis Test when σ_1 and σ_2 are unknown

$$t = \frac{(\overline{x}_1 - \overline{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$$

with $d.f.$ = smaller of $n_1 - 1$ and $n_2 - 1$

Matched pairs (dependent samples)

Confidence Interval

$$\overline{d} - E < \mu_{\overline{d}} < \overline{d} + E$$

where $E = t_{\alpha/2}\dfrac{s_d}{\sqrt{n}}$ with d.f. $= n - 1$

Hypothesis Test

$$t = \frac{\overline{d} - \mu_{\overline{d}}}{\dfrac{s_d}{\sqrt{n}}} \text{ with } d.f. = n - 1$$

Two Sample Variances

Confidence Interval for σ_1^2 and σ_2^2

$$\left(\frac{s_1^2}{s_2^2} \bullet \frac{1}{F_{right}}\right) < \frac{\sigma_1^2}{\sigma_2^2} < \left(\frac{s_1^2}{s_2^2} \bullet \frac{1}{F_{left}}\right)$$

Hypothesis Test Statistic: $F = \dfrac{s_1^2}{s_2^2}$ where $s_1^2 \ge s_2^2$

numerator $d.f. = n_1 - 1$ and denominator $d.f. = n_2 - 1$

Chapter 10

Regression and Correlation

Linear Correlation Coefficient (r)

$$r = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{n(\sum x^2) - (\sum x)^2}\sqrt{n(\sum y^2) - (\sum y)^2}}$$

OR

$$r = \frac{\Sigma(z_x z_y)}{n-1} \quad \text{where } z_x = \text{ z score for x and } z_y = \text{ z score for y}$$

Coefficient of Determination: $r^2 = \dfrac{\text{explained variation}}{\text{total variation}}$

Standard Error of Estimate: $s_e = \sqrt{\dfrac{\Sigma(y-\hat{y})^2}{n-2}}$

or $s_e = \sqrt{\dfrac{\sum y^2 - b_0 \sum y - b_1 \sum xy}{n-2}}$

Prediction Interval: $\hat{y} - E < y < \hat{y} + E$

where $E = t_{\alpha/2}\ s_e \sqrt{1 + \dfrac{1}{n} + \dfrac{n(x_0 - \bar{x})^2}{n(\Sigma x^2) - (\Sigma x)^2}}$

Sample test statistic for r

$$t = \frac{r}{\sqrt{\dfrac{1-r^2}{n-2}}} \quad \text{with } d.f. = n-2$$

Least-Squares Line (Regression Line or Line of Best Fit)

$\hat{y} = b_0 + b_1 x$ note that b_0 is the y-intercept and b_1 is the slope

where $b_1 = \dfrac{n\sum xy - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2}$ or $b_1 = r\dfrac{s_y}{s_x}$

and

where $b_0 = \dfrac{(\sum y)(\sum x^2) - (\sum x)(\sum xy)}{n(\sum x^2) - (\sum x)^2}$ or $b_0 = \bar{y} - b_1\bar{x}$

Confidence interval for y-intercept β_0

$b_0 - E < \beta_0 < b_0 + E$

where E = $t_{\alpha/2}\ s_e \sqrt{\dfrac{1}{n} + \dfrac{\bar{x}^2}{\sum x^2 - \dfrac{(\sum x)^2}{n}}}$

Confidence interval for slope β_1

$b_1 - E < \beta_1 < b_1 + E$

where E = $t_{\alpha/2} \bullet \dfrac{s_e}{\sqrt{\sum x^2 - \dfrac{(\sum x)^2}{n}}}$

Chapter 11

$$\chi^2 = \sum \frac{(O-E)^2}{E} \quad \text{where } E = \frac{\text{(row total)(column total)}}{\text{sample size}}$$

Tests of Independence $d.f. = (R-1)(C-1)$

Goodness of fit $d.f. = (\text{number of categories}) - 1$

Chapter 12

One Way ANOVA

$k =$ number of groups; $N =$ total sample size

$$SS_{TOT} = \sum x_{TOT}^2 - \frac{(\sum x_{TOT})^2}{N}$$

$$SS_{BET} = \sum_{\text{all groups}} \left(\frac{(\sum x_i)^2}{n_i} \right) - \frac{(\sum x_{TOT})^2}{N}$$

$$SS_W = \sum_{\text{all groups}} \left(\sum x_i^2 - \frac{(\sum x_i)^2}{n_i} \right)$$

$$SS_{TOT} = SS_{BET} + SS_W$$

$$MS_{BET} = \frac{SS_{BET}}{d.f._{BET}} \quad \text{where } d.f._{BET} = k-1$$

$$MS_W = \frac{SS_W}{d.f._W} \quad \text{where } d.f._W = N-k$$

$F = \dfrac{MS_{BET}}{MS_W}$ where $d.f.$ numerator = $d.f._{BET} = k-1$

$d.f.$ denominator = $d.f._W = N-k$

Two-Way ANOVA

r = number of rows; c = number of columns

Row factor F : $\dfrac{MS \text{ row factor}}{MS \text{ error}}$

Column factor F : $\dfrac{MS \text{ column factor}}{MS \text{ error}}$

Interaction F : $\dfrac{MS \text{ interaction}}{MS \text{ error}}$

with degrees of freedom for

row factor = $r-1$

column factor = $c-1$

interaction = $(r-1)(c-1)$

error = $rc(n-1)$

NEGATIVE z Scores

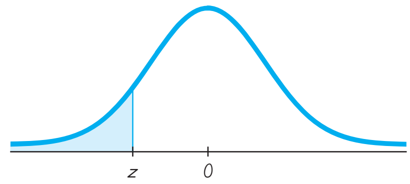


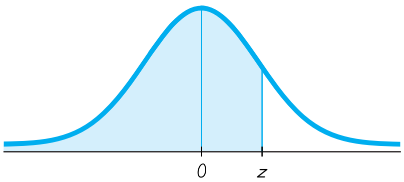
TABLE A-2 Standard Normal (z) Distribution: Cumulative Area from the LEFT

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.50 and lower	.0001									
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	* .0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	* .0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
−0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
−0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
−0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
−0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
−0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
−0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
−0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
−0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
−0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

NOTE: For values of z below −3.49, use 0.0001 for the area.

*Use these common values that result from interpolation:

z score	Area
−1.645	0.0500
−2.575	0.0050



POSITIVE *z* Scores

TABLE A-2 (*continued*) Cumulative Area from the LEFT

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495 *	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949 *	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998
3.50 and up	.9999									

NOTE: For values of *z* above 3.49, use 0.9999 for the area.

*Use these common values that result from interpolation:

z score	Area
1.645	0.9500
2.575	0.9950

Common Critical Values

Confidence Level	Critical Value
0.90	1.645
0.95	1.96
0.99	2.575

critical z-values for hypothesis testing

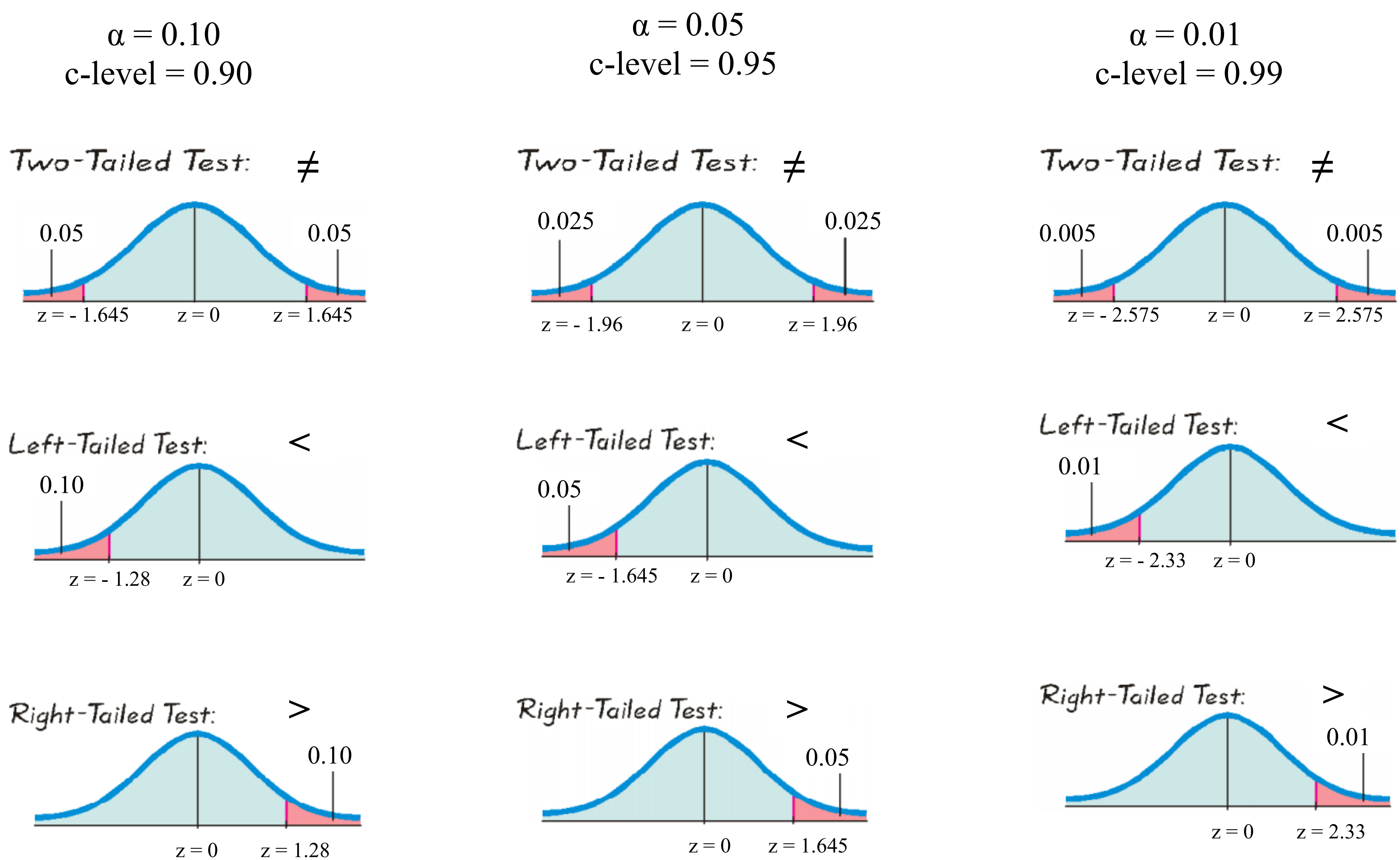


Figure 8.4

TABLE A-3	*t* Distribution: Critical *t* Values				
	Area in One Tail				
	0.005	0.01	0.025	0.05	0.10
	Area in Two Tails				
Degrees of Freedom	0.01	0.02	0.05	0.10	0.20
1	**63.657**	31.821	**12.706**	6.314	3.078
2	**9.925**	6.965	**4.303**	2.920	1.886
3	**5.841**	4.541	**3.182**	2.353	1.638
4	**4.604**	3.747	**2.776**	2.132	1.533
5	**4.032**	3.365	**2.571**	2.015	1.476
6	**3.707**	3.143	**2.447**	1.943	1.440
7	**3.499**	2.998	**2.365**	1.895	1.415
8	**3.355**	2.896	**2.306**	1.860	1.397
9	**3.250**	2.821	**2.262**	1.833	1.383
10	**3.169**	2.764	**2.228**	1.812	1.372
11	**3.106**	2.718	**2.201**	1.796	1.363
12	**3.055**	2.681	**2.179**	1.782	1.356
13	**3.012**	2.650	**2.160**	1.771	1.350
14	**2.977**	2.624	**2.145**	1.761	1.345
15	**2.947**	2.602	**2.131**	1.753	1.341
16	**2.921**	2.583	**2.120**	1.746	1.337
17	**2.898**	2.567	**2.110**	1.740	1.333
18	**2.878**	2.552	**2.101**	1.734	1.330
19	**2.861**	2.539	**2.093**	1.729	1.328
20	**2.845**	2.528	**2.086**	1.725	1.325
21	**2.831**	2.518	**2.080**	1.721	1.323
22	**2.819**	2.508	**2.074**	1.717	1.321
23	**2.807**	2.500	**2.069**	1.714	1.319
24	**2.797**	2.492	**2.064**	1.711	1.318
25	**2.787**	2.485	**2.060**	1.708	1.316
26	**2.779**	2.479	**2.056**	1.706	1.315
27	**2.771**	2.473	**2.052**	1.703	1.314
28	**2.763**	2.467	**2.048**	1.701	1.313
29	**2.756**	2.462	**2.045**	1.699	1.311
30	**2.750**	2.457	**2.042**	1.697	1.310
31	**2.744**	2.453	**2.040**	1.696	1.309
32	**2.738**	2.449	**2.037**	1.694	1.309
33	**2.733**	2.445	**2.035**	1.692	1.308
34	**2.728**	2.441	**2.032**	1.691	1.307
35	**2.724**	2.438	**2.030**	1.690	1.306
36	**2.719**	2.434	**2.028**	1.688	1.306
37	**2.715**	2.431	**2.026**	1.687	1.305
38	**2.712**	2.429	**2.024**	1.686	1.304
39	**2.708**	2.426	**2.023**	1.685	1.304
40	**2.704**	2.423	**2.021**	1.684	1.303
45	**2.690**	2.412	**2.014**	1.679	1.301
50	**2.678**	2.403	**2.009**	1.676	1.299
60	**2.660**	2.390	**2.000**	1.671	1.296
70	**2.648**	2.381	**1.994**	1.667	1.294
80	**2.639**	2.374	**1.990**	1.664	1.292
90	**2.632**	2.368	**1.987**	1.662	1.291
100	**2.626**	2.364	**1.984**	1.660	1.290
200	**2.601**	2.345	**1.972**	1.653	1.286
300	**2.592**	2.339	**1.968**	1.650	1.284
400	**2.588**	2.336	**1.966**	1.649	1.284
500	**2.586**	2.334	**1.965**	1.648	1.283
1000	**2.581**	2.330	**1.962**	1.646	1.282
2000	**2.578**	2.328	**1.961**	1.646	1.282
Large	**2.576**	2.326	**1.960**	1.645	1.282

TABLE A-4 Chi-Square (χ^2) Distribution

Degrees of Freedom	Area to the Right of the Critical Value									
	0.995	0.99	0.975	0.95	0.90	0.10	0.05	0.025	0.01	0.005
1	—	—	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.071	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.299
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.042	30.813	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	18.114	36.741	40.113	43.194	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.121	14.257	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.336
30	13.787	14.954	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
40	20.707	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766
50	27.991	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490
60	35.534	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952
70	43.275	45.442	48.758	51.739	55.329	85.527	90.531	95.023	100.425	104.215
80	51.172	53.540	57.153	60.391	64.278	96.578	101.879	106.629	112.329	116.321
90	59.196	61.754	65.647	69.126	73.291	107.565	113.145	118.136	124.116	128.299
100	67.328	70.065	74.222	77.929	82.358	118.498	124.342	129.561	135.807	140.169

From Donald B. Owen, *Handbook of Statistical Tables,* © 1962 Addison-Wesley Publishing Co., Reading, MA. Reprinted with permission of the publisher.

Degrees of Freedom

$n - 1$	for confidence intervals or hypothesis tests with a standard deviation or variance
$k - 1$	for goodness-of-fit with k categories
$(r - 1)(c - 1)$	for contingency tables with r rows and c columns
$k - 1$	for Kruskal-Wallis test with k samples

Determining Sample Size for Population Variance or Standard Deviation

Table 7-2

Sample Size for σ^2		Sample Size for σ	
To be 95% confident that s^2 is within	of the value of σ^2, the sample size n should be at least	To be 95% confident that s is within	of the value of σ, the sample size n should be at least
1%	77,208	1%	19,205
5%	3,149	5%	768
10%	806	10%	192
20%	211	20%	48
30%	98	30%	21
40%	57	40%	12
50%	38	50%	8
To be 99% confident that s^2 is within	of the value of σ^2, the sample size n should be at least	To be 99% confident that s is within	of the value of σ, the sample size n should be at least
1%	133,449	1%	33,218
5%	5,458	5%	1,336
10%	1,402	10%	336
20%	369	20%	85
30%	172	30%	38
40%	101	40%	22
50%	68	50%	14

(table 7-2 from page 390, Triola 4th edition)

TABLE A-6
Critical Values of the
Pearson Correlation Coefficient r

n	alpha = .05	alpha = .01
4	0.950	0.990
5	0.878	0.959
6	0.811	0.917
7	0.754	0.875
8	0.707	0.834
9	0.666	0.798
10	0.632	0.765
11	0.602	0.735
12	0.576	0.708
13	0.553	0.684
14	0.532	0.661
15	0.514	0.641
16	0.497	0.623
17	0.482	0.606
18	0.468	0.590
19	0.456	0.575
20	0.444	0.561
25	0.396	0.505
30	0.361	0.463
35	0.335	0.430
40	0.312	0.402
45	0.294	0.378
50	0.279	0.361
60	0.254	0.330
70	0.236	0.305
80	0.220	0.286
90	0.207	0.269
100	0.196	0.256

NOTE: To test H0: $\rho = 0$ against H1: $\rho \neq 0$, reject H0 if the absolute value of r is greater than the critical value in the table.

Greek Alphabet

Greek Letter	Name	Equivalent	Sound When Spoken
Α α	Alpha	A	al-fah
Β β	Beta	B	bay-tah
Γ γ	Gamma	G	gam-ah
Δ δ	Delta	D	del-tah
Ε ε	Epsilon	E	ep-si-lon
Ζ ζ	Zeta	Z	zay-tah
Η η	Eta	E	ay-tay
Θ θ	Theta	Th	thay-tah
Ι ι	Iota	I	eye-o-tah
Κ κ	Kappa	K	cap-ah
Λ λ	Lambda	L	lamb-dah
Μ μ	Mu	M	mew
Ν ν	Nu	N	new
Ξ ξ	Xi	X	zzEye
Ο ο	Omicron	O	om-ah-cron
Π π	Pi	P	pie
Ρ ρ	Rho	R	row
Σ σ	Sigma	S	sig-ma
Τ τ	Tau	T	tawh
Υ υ	Upsilon	U	oop-si-lon
Φ φ	Phi	Ph	figh or fie
Χ χ	Chi	Ch	kigh
Ψ ψ	Psi	Ps	sigh
Ω ω	Omega	O	o-may-gah

http://www.keyway.ca/htm2002/greekal.htm

Summary of Formulas and Concepts

Descriptive Statistics (Ch. 1-4)

Definitions

Population: The complete set of numerical information on a particular quantity in which an investigator is interested. We assume a population consists of N values.

Sample: An observed subset of population values. We assume a sample consists of n values.

Graphic Summaries

Dotplot - A graphic to display the original data. Draw a number line, and put a dot on it for each observation. For identical or really close observations, stack the dots.

```
                             .
                         .   :  .:  .:  :
                  .    . :  .:..::  ::  :.   .
       ..   . : . ::.: :.:.:::::::::::..::::::. ...  : .
 +---------+---------+---------+---------+---------+-------
40        50        60        70        80        90
```

Histogram - A graphic that displays the shape of numeric data by grouping it in intervals.

1. Choose evenly spaced categories
2. Count the number of observations in each group
3. Draw a bar graph with the height of each bar equal to the number of observations in the corresponding interval.

Stem and leaf plot Similar to a dotplot. Data are grouped according to their leading digits, and the last digit is used as a plotting symbol (like a dot in the dotplot).

The left digits are a cumulative count on each side of the middle. The bracketed number is how many observations are in the middle. The middle column of digits are the first digits of the number, and the "bars" are the last digit.

```
   2    4 57
   5    5 133
  13    5 56777899
  23    6 0222334444
  39    6 5667777788888899
(18)    7 000000111222222334
  38    7 5555555666668899999999
  16    8 0022333444
```

Numeric Summaries of Data

Suppose that we have n observations, labeled $x_1, x_2, \ldots, x_n$.
Then $\sum_{i=1}^n x_i$ means

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + x_3 + \ldots + x_n$$

Some other relations are:

$\displaystyle\sum_{i=1}^{n} f(x_i) = f(x_1) + f(x_2) + f(x_3) + \ldots + f(x_n)$, for any function f,

$\displaystyle\sum_{i=1}^{n} cx_i = c\sum_{i=1}^{n} x_i$ for any constant c,

$\displaystyle\sum_{i=1}^{n}(ax_i + by_i) = a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} y_i$, for any constants a and b.

Measures of Location - numeric summaries of the center of a distribution.

Mean (or average)

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i \qquad \text{(population)} \qquad \overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \text{(sample)}$$

Median - the middle observation

The middle observation of the sorted data if n is odd, otherwise the average of the two middle values.

Measures of Dispersion - numeric summaries of the spread or variation of a distribution.

Variance and Standard Deviation

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2 \qquad \text{(population)} \qquad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

Population and sample standard deviations (σ and s) are just the square roots of these quantities.

Interpreting σ

Chebyshev's rule: for *any* population
at least 75% of the observations lie within 2σ of μ,
at least 89% of the observations lie within 3σ of μ,
at least $100(1 - 1/m^2)\%$ of the observations lie within $m \times \sigma$ of the mean μ.

IQR (Interquartile Range): The distance between the $(n+1)/4$th and $3 \times (n+1)/4$th observations in an ordered dataset. These two values are called the first and third quartiles.

Measure of symmetry : Skewness

$$\text{skewness} = \sum_{i=1}^{n} \frac{(x_i - \overline{x})^3/n}{s^3}$$

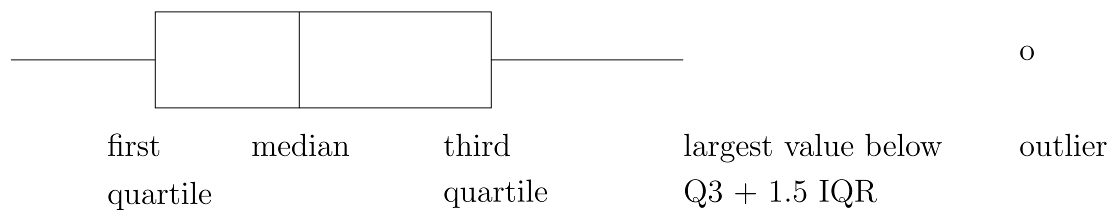
Negative means left skewed, 0 means symmetric, positive means right skewed.

Measure of heavy tails : Kurtosis

$$\text{kurtosis} = \sum_{i=1}^{n} \frac{(x_i - \overline{x})^4/n}{s^4} - 3$$

A normal distribution has a kurtosis of 3, so if we subtract 3, interpretations are relative to that. Positive values mean a sharper peak, and negative values mean a flatter top than a normal distribution.

Box-and-whisker plot: A graphic that summarizes the data using the median and quartiles, and displays outliers. Good for comparing several groups of data



Probability (Ch. 5)

Definitions and Set Theory:

Random experiment: A process leading to at least two possible outcomes with uncertainty as to which will occur.

Basic outcome: A possible outcome of the random experiment.

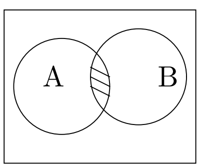
Sample space: The set of all basic outcomes.

Event: A set of basic outcomes from the sample space. An event is said to occur if any one of its constituent basic outcomes occurs.

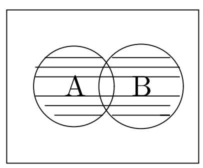
Combining events: let A and B be two events.

Technical	Symbol	Pronounced	Meaning
Union	$A \cup B$	A or B	A occurs or B occurs or both occur
Intersection	$A \cap B$	A and B	A occurs and B occurs
Complement	$\overline{A}$	not A	A does not occur

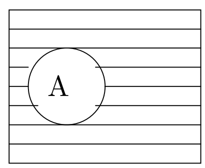
Venn Diagrams:



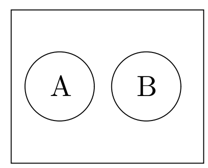
shaded = A ∩ B



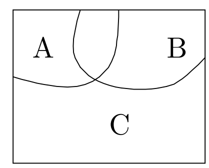
shaded = A ∪ B



$\overline{A}$ shaded



A, B mutually exclusive



A,B,C collectively exhaustive

Probability Postulates:

1. If A is any event in the sample space S, $0 \leq \mathrm{P}(A) \leq 1$
2. Let A be an event in S and let O_i denote the basic outcomes. Then $\mathrm{P}(A) = \sum_A \mathrm{P}(O_i)$, where the notation implies that the summation extends over all the basic outcomes in A.
3. $\mathrm{P}(S) = 1$

Probability rules for combining events:

$\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cap B)$

$\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B)$ if A and B mutually exclusive.

$\mathrm{P}(\overline{A}) = 1 - \mathrm{P}(A)$

Conditional Probability:

$\mathrm{P}(A|B) = \dfrac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)}$ provided $\mathrm{P}(B) > 0$.

$\mathrm{P}(A \cap B) = \mathrm{P}(A|B)\mathrm{P}(B) = \mathrm{P}(B|A)\mathrm{P}(A)$

Independence:

Two events are *Statistically Independent* if and only if

$$P(A \cap B) = P(A)P(B)$$

or equivalently $P(A|B) = P(A)$ and $P(B|A) = P(B)$.
General case: events $E_1, E_2, \ldots, E_k$ are independent if and only if

$$P(E_1 \cap E_2 \cap \ldots \cap E_k) = P(E_1)P(E_2)\ldots P(E_k)$$

Bivariate Probability

Probabilities of outcomes for bivariate events:

	B_1	B_2	B_3	
A_1	$P(A_1 \cap B_1)$	$P(A_1 \cap B_2)$	$P(A_1 \cap B_3)$	$P(A_1)$
A_2	$P(A_2 \cap B_1)$	$P(A_2 \cap B_2)$	$P(A_2 \cap B_3)$	$P(A_2)$
	$P(B_1)$	$P(B_2)$	$P(B_3)$	

$P(A_1 \cap B_2)$ is a probability of A_1 and B_2 occurring.

$P(A_1) = P(A_1 \cap B_1) + P(A_1 \cap B_2) + P(A_1 \cap B_3)$ is the **marginal probability** that A_1 occurs.

If we think of A_1, A_2 as a group of attributes A, and B_1, B_2, B_3 as a group of attributes B, then A **and** B **are independent** only if every one of $\{A_1, A_2\}$ are independent of every one of $\{B_1, B_2, B_3\}$.

Discrete Random Variables (Ch. 6-7)

Definitions

Random Variable: (r.v.) A variable that takes on numerical values determined by the outcome of a random experiment.

Discrete Random Variable: A r.v. that can take on no more than a countable number of values.

Continuous Random Variable: A r.v. that can take any value in an interval.

Notation: An upper case letter (e.g. X) will represent a r.v.; a lower case letter (e.g. x) will represent one of its possible values.

Discrete Probability Distributions

The **probability function**, $\mathrm{P}_X(x)$, of a discrete r.v. X gives the probability that X takes the value x:

$$\mathrm{P}_X(x) = \mathrm{P}(X = x)$$

where the function is evaluated at all possible values of x.

Properties:

1. $\mathrm{P}_X(x) \geq 0$ for any value x
2. $\sum_x \mathrm{P}_X(x) = 1$

Cumulative probability function, $F_X(x_0)$ of a r.v. X:

$$F_X(x_0) = P(X \leq x_0) = \sum_{x \leq x_0} \mathrm{P}_X(x).$$

Properties:

1. $0 \leq F_X(x) \leq 1$ for any x
2. If $a < b$, then $F_X(a) \leq F_X(b)$.

Expectation of Discrete Random Variables

Expected value of a discrete r.v.:

$$E(X) = \mu_X = \sum_x x P_X(x).$$

$$\text{For any function } g(X), \qquad E(g(X)) = \sum_x g(x) P_X(x).$$

Variance of a discrete r.v.:

$$\mathrm{Var}(X) = \sigma_X^2 = E((X - \mu_X)^2) = \sum_x (x - \mu_X)^2 \mathrm{P}_X(x) = E(X^2) - \mu_X^2.$$

The **standard deviation** of X is σ_X.

Plug-In Rules: let X be a r.v., and a and b constants. Then

$$E(a + bX) = a + bE(X)$$

$$\mathrm{Var}(a + bX) = b^2 \mathrm{Var}(X).$$

This only works for linear functions.

Jointly Distributed Discrete Random Variables

Joint Probability Function: Suppose X and Y are r.v.'s. Their joint probability function gives the probability that simultaneously $X = x$ and $Y = y$:

$$\mathrm{P}_{X,Y}(x,y) = \mathrm{P}(\{X = x\} \cap \{Y = y\})$$

Properties:

1. $\mathrm{P}_{X,Y}(x,y) \geq 0$ for any pair (x,y)
2. $\sum_x \sum_y \mathrm{P}_{X,Y}(x,y) = 1$.

Marginal probability function:

$$\mathrm{P}_X(x) = \sum_y \mathrm{P}_{X,Y}(x,y).$$

Conditional probability function:

$$\mathrm{P}_{Y|X}(y|x) = \frac{\mathrm{P}_{X,Y}(x,y)}{\mathrm{P}_X(x)}$$

Independence: X and Y are independent if and only if

$$\mathrm{P}_{X,Y}(x,y) = \mathrm{P}_X(x)\mathrm{P}_Y(y) \qquad \text{for all possible } (x,y) \text{ pairs}$$

Expectation: Let X and Y be r.v.'s, and $g(X,Y)$ any function. Then

$$E(g(X,Y)) = \sum_x \sum_y g(x,y)\mathrm{P}_{X,Y}(x,y).$$

Conditional Expectation: Let X and Y be r.v.'s, and suppose we know the conditional distribution of X for $Y = y$, labeled $\mathrm{P}_{X|Y}(x|y)$. Then

$$\mathrm{E}(X|Y = y) = \sum_x x\mathrm{P}_{X|Y}(x|y).$$

Covariance: If $E(X) = \mu_X$ and $E(Y) = \mu_Y$,

$$\mathrm{Cov}(X,Y) = E[(X - \mu_X)(Y - \mu_Y)] = \sum_x \sum_y (x - \mu_X)(y - \mu_Y)\mathrm{P}_{X,Y}(x,y)$$

$$= E(XY) - \mu_X\mu_Y = \left[\sum_x \sum_y xy\mathrm{P}_{X,Y}(x,y)\right] - \mu_X\mu_Y$$

If two r.v.'s are independent, their covariance is zero. The converse is not necessarily true.

Correlation:

$$\rho_{XY} = \mathrm{Cor}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}$$

$-1 \leq \rho_{XY} \leq 1$ always.
$\rho_{XY} = \pm 1$ if and only if $Y = a + bX$ (a,b constants).

Plug-in rules: Let X and Y be r.v.'s, and a, b constants. Then

$$E(aX + bY) = aE(X) + bE(Y)$$

$$\mathrm{Var}(aX + bY) = a^2\mathrm{Var}(X) + b^2\mathrm{Var}(Y) + 2ab\mathrm{Cov}(X,Y)$$

Binomial distribution

Bernoulli Trials:
A sequence of repeated experiments are Bernoulli trials if:

1. The result of each trial is either a success or failure.
2. The probability p of a success is the same for all trials.
3. The trials are *independent*.

If X is the number of successes in n Bernoulli trials, X is a **Binomial Random Variable**. It has probability function:

$$\mathrm{P}_X(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

Where $\binom{n}{x}$ counts the number of ways of getting x successes in n trials. The formula for $\binom{n}{x}$ is

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}$$

where $n! = n \times (n-1) \times (n-2) \times ... \times 2 \times 1$.

Mean and Variance: $E(X) = np$, $\mathrm{Var}(X) = np(1-p)$.

Continuous Random Variables (Ch. 7)

Probability Distributions

Probability density function: A function $f_X(x)$ of the continuous r.v. X with the following properties:

1. $f_X(x) \geq 0$ for all values of x.
2. $\mathrm{P}(a \leq X \leq b) =$ the area under $f_X(x)$ between a and b, if $a < b$.
3. The total area under the curve is 1
4. The area under the curve to the left of any value x is $F_X(x)$, the probability that X does not exceed x.

Cumulative distribution function: Same as before.

$$\mathrm{P}(a \leq X \leq b) = F_X(b) - F_X(a) \qquad \text{(provided } a < b\text{).}$$

Expectations, Variances, Covariances, etc.
Same rules as for discrete r.v.'s. The summation ($\sum$) is replaced by the integral ($\int$), which is not necessary for this course.

Normal Distribution

Probability Density function:

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}$$

for constants μ and σ such that $-\infty < \mu < \infty$ and $0 < \sigma < \infty$.
Mean and Variance: $E(X) = \mu \qquad \mathrm{Var}(X) = \sigma^2$
Notation: $X \sim N(\mu, \sigma^2)$ means X is normal with mean μ and variance σ^2.

If $Z \sim N(0,1)$ we say it has a **standard normal distribution**.
If $X \sim N(\mu, \sigma^2)$ then $Z = (X - \mu)/\sigma \sim N(0,1)$. Thus

$$\mathrm{P}(a < X < b) = \mathrm{P}\left(\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\right) = F_Z\left(\frac{b-\mu}{\sigma}\right) - F_Z\left(\frac{a-\mu}{\sigma}\right)$$

Central Limit Theorem

Let $X_1, X_2, \ldots, X_n$ be n independent r.v.'s, each with identical distributions, mean μ and variance σ^2. As n becomes large,

$$\overline{X} \sim N(\mu, \sigma^2/n)$$

$$\sum_{i=1}^{n} X_i = n\overline{X} \sim N(n\mu, n\sigma^2)$$

Sampling & Sampling distributions

Simple random sample: (or random sample) A method of randomly drawing n objects wihich are **Independent and Identically Distributed (I.I.D.)**.

Statistic: A function of the sample information.

Sampling distribution of a statistic: The probability distribution of the values a statistic can take, over all possible samples of a fixed size n.

Sampling distribution of the mean : Suppose $X_1, \ldots X_n$ are a random sample from some population with mean μ_X and variance σ_X^2. The **sample mean** is

$$\overline{X} = \sum_{i=1}^{n} X_i \ .$$

It has the following properties:

1. $\mathrm{E}(\overline{X}) = \mu_X$
2. It has standard deviation $\sigma_{\overline{X}} = \sigma_X / \sqrt{n}$.
3. If the population distribution is normal,

$$\overline{X} \sim N(\mu_X, \sigma_{\overline{X}}^2) = N(\mu_X, \sigma_X^2/n).$$

4. If the population distribution is not normal, but n is large, then (3) is roughly true.

Sampling distribution of a proportion : Suppose the r.v. X is the number of successes in a binomial sample of n trials, whose probability of success is p. The **sample proportion** is

$$\hat{p} = X/n$$

It has the following properties:

1. $\mathrm{E}(\hat{p}) = p$
2. It has standard deviation $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$.
3. If n is large ($np(1-p) > 9$ or roughly $n \geq 40$),

$$\hat{p} \sim N(p, \sigma_{\hat{p}}^2) = N(p, p(1-p)/n).$$

Point Estimation

Estimator: A random variable that depends on the sample information and whose realizations provide approximations to an unknown population parameter.

Estimate: A specific realization of an estimator.

Point estimator: An estimator that is a single number.

Point estimate: A specific realization of a point estimator.

Bias: Let $\hat{\theta}$ be an estimate of the parameter θ. The bias in $\hat{\theta}$ is

$$\mathrm{Bias}(\hat{\theta}) = \mathrm{E}(\hat{\theta}) - \theta.$$

If the bias is 0, $\hat{\theta}$ is an **unbiased estimator**.

Efficiency: Let $\hat{\theta}_1$ and $\hat{\theta}_2$ be two estimators of θ, based on the same sample. Then θ_1 is **more efficient** than $\hat{\theta}_2$ if

$$\mathrm{Var}(\hat{\theta}_1) < \mathrm{Var}(\hat{\theta}_2).$$

Interval Estimation

Confidence Interval: Let θ be an unknown parameter. Suppose that from sample information, we can find random variables A and B such that

$$\mathrm{P}(A < \theta < B) = 1 - \alpha.$$

If the observed values are a and b, then (a, b) is a $100(1-\alpha)\%$ confidence interval for θ. The quantity $(1 - \alpha)$ is called the *probability content* of the interval.

Student's t distribution: Given a random sample of n observations with mean $\overline{X}$ and standard deviation s, from a normal population with mean μ, the random variable

$$T = \frac{\overline{X} - \mu}{s/\sqrt{n}}$$

follows the Student's t distribution with $(n-1)$ degrees of freedom. For $n > 30$, the t distribution is quite close to a $N(0, 1)$ distribution.

Data	Parameter	$100(1-\alpha)\%$ C.I.
$N(\mu, \sigma^2)$, σ^2 known	μ	$\overline{x} \pm z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}$
mean μ, σ^2 unknown, $n > 30$	μ	$\overline{x} \pm z_{\alpha/2}\dfrac{s}{\sqrt{n}}$
$N(\mu, \sigma^2)$, σ^2 unknown	μ	$\overline{x} \pm t_{n-1,\alpha/2}\dfrac{s}{\sqrt{n}}$
Binomial(n, p), $np(1-p) > 9$, or roughly $n \geq 40$	p	$\hat{p} \pm z_{\alpha/2}\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$
n matched pairs, difference $\sim N(\mu_X - \mu_Y, \sigma^2)$	$\mu_X - \mu_Y$	$\overline{d} \pm t_{n-1,\alpha/2}\dfrac{s_d}{\sqrt{n}}$
2 independent samples, means μ_X, μ_Y variances unknown, $n > 30$	$\mu_X - \mu_Y$	$\overline{x} - \overline{y} \pm z_{\alpha/2}\sqrt{\dfrac{s_x^2}{n_x} + \dfrac{s_y^2}{n_y}}$
2 independent samples, means μ_X, μ_Y variances unknown	$\mu_X - \mu_Y$	$\overline{x} - \overline{y} \pm t_{n^*,\alpha/2}\sqrt{\dfrac{s_x^2}{n_x} + \dfrac{s_y^2}{n_y}}$
2 independent samples, Binomial(n_X, p_X),Binomial(n_Y, p_Y)	$p_X - p_Y$	$\hat{p}_x - \hat{p}_y \pm z_{\alpha/2}\sqrt{\dfrac{\hat{p}_x(1-\hat{p}_x)}{n_x} + \dfrac{\hat{p}_y(1-\hat{p}_y)}{n_y}}$

Notes for the table:

1. All quantities in the C.I. column are either known constants or observed sample quantities.
2. $\text{P}(Z > z_{\alpha/2}) = \alpha/2$ for $Z \sim N(0, 1)$.
3. $\text{P}(T > t_{n-1,\alpha/2}) = \alpha/2$ for $T \sim$ Student's t with $(n-1)$ d.f.
4. s, s_x, s_y, s_d are observed sample standard deviations corresponding to $x_i, x_i, y_i, d_i = x_i - y_i$ respectively.
5. $\hat{p}, \hat{p}_x, \hat{p}_y$ are the observed sample proportions corresponding to x_i, x_i, y_i respectively.
6. $\overline{d}$ is the sample mean corresponding to $d_i = x_i - y_i$.
7. n, n_x, n_y are the total, x and y sample sizes.

8. $n^* = \left(\dfrac{s_x^2}{n_x} + \dfrac{s_y^2}{n_y}\right)^2 \Bigg/ \left[\dfrac{(s_x^2/n_x)^2}{n_x - 1} + \dfrac{(s_y^2/n_y)^2}{n_y - 1}\right]$

Also note: The text gives a different formula for comparing two means with small sample sizes. It requires that the two variances be the same, which may not be the case. Unless you're sure the variances are equal, it's safer to use the approximation given here (the formula with a n^*). If you are sure that the variances are equal, using the book's formula is ok.

Estimating the sample size: If you want an $100(1-\alpha)\%$ interval of $\pm L$ (i.e. length $2L$), choose n so

situation	n
normal, σ known	$n = \dfrac{z_{\alpha/2}^2\sigma^2}{L^2}$
Bernoulli, worst case	$n = \dfrac{0.25z_{\alpha/2}^2}{L^2}$

Hypothesis Testing

Null Hypothesis (H_0): The hypothesis we assume to be true unless there is sufficient evidence to the contrary.

Alternative Hypothesis (H_1): The hypothesis we test the null against. If there is evidence that H_0 is false, we accept H_1.

Type I Error: Rejecting a true H_0.

Type II Error: Not rejecting a false H_0.

Significance Level: P(reject$H_0|H_0$true) = P(type I error).

Power: The probability of rejecting a null hypothesis that is false. Note that this depends on the true value of the parameter.

P-value: The smallest significance level at which a null hypothesis can be rejected. This is a measure of how likely the data is, if H_0 is true.

Notes for the following table (In addition to the comments for CI's):

1. The first three tests are examples of one ($>$ and $<$ alternatives) and two sided ($\neq$ alternative) tests. The remaining tests all have a $>$ alternative, but are easily adaptable to either of the other two alternatives.

2. In the last test,

$$\hat{p}_0 = \frac{n_x \hat{p}_x + n_y \hat{p}_y}{n_x + n_y}$$

3. In all the tests, we are comparing the unknown parameter (such as μ, p or $\mu_X - \mu_Y$) to constants (μ_0, p_0, and D_0).

Hypothesis tests with significance level α

Data	H_0	H_1	reject H_0 if
$N(\mu, \sigma^2)$, σ^2 known	$\mu = \mu_0$ or $\mu \le \mu_0$	$\mu > \mu_0$	$\frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}} > z_\alpha$
same	$\mu = \mu_0$ or $\mu \ge \mu_0$	$\mu < \mu_0$	$\frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}} < -z_\alpha$
same	$\mu = \mu_0$	$\mu \neq \mu_0$	$\frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}}$ not in $(-z_{\alpha/2}, z_{\alpha/2})$
mean μ, σ^2 unknown $n > 30$	$\mu = \mu_0$ or $\mu \le \mu_0$	$\mu > \mu_0$	$\frac{\overline{x} - \mu_0}{s/\sqrt{n}} > z_\alpha$
$N(\mu, \sigma^2)$, σ^2 unknown $n < 30$	$\mu = \mu_0$ or $\mu \le \mu_0$	$\mu > \mu_0$	$\frac{\overline{x} - \mu_0}{s/\sqrt{n}} > t_{n-1,\alpha}$
Binomial(n, p) $n(1-p) > 9$ or roughly $n >$ 40	$p = p_0$ or $p \le p_0$	$p > p_0$	$\frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} > z_\alpha$
n matched pairs, difference $\sim N(\mu_X - \mu_Y, \sigma^2)$	$\mu_X - \mu_Y = D_0$ or $\mu_X - \mu_Y \le D_0$	$\mu_X - \mu_Y > D_0$	$\frac{\overline{d} - D_0}{s_d/\sqrt{n}} > t_{n-1,\alpha}$

(Continued on next page)

Hypothesis tests with significance level α

Data	H_0	H_1	reject H_0 if
2 independent samples, means μ_X, μ_Y variances unknown, $n_x > 30, n_y > 30$	$\mu_X - \mu_Y = D_0$ or $\mu_X - \mu_Y \leq D_0$	$\mu_X - \mu_Y > D_0$	$\dfrac{\overline{x} - \overline{y} - D_0}{\sqrt{\dfrac{s_x^2}{n_x} + \dfrac{s_y^2}{n_y}}} > z_\alpha$
2 independent normal samples, means μ_X, μ_Y variances unknown	$\mu_X - \mu_Y = D_0$ or $\mu_X - \mu_Y \leq D_0$	$\mu_X - \mu_Y > D_0$	$\dfrac{\overline{x} - \overline{y}}{\sqrt{\dfrac{s_x^2}{n_x} + \dfrac{s_y^2}{n_y}}} > t_{n^*,\alpha}$
2 independent samples, Binomial(n_x, p_x) and Binomial(n_y, p_y)	$p_x - p_y = 0$ or $p_x - p_y \leq 0$	$p_x - p_y > 0$	$\dfrac{\hat{p}_x - \hat{p}_y}{\sqrt{\hat{p}_0(1-\hat{p}_0)\left(\dfrac{n_x + n_y}{n_x n_y}\right)}} > z_\alpha$

AP Statistics Formula Sheet

(I) Descriptive Statistics

$$\bar{x} = \frac{\sum x_i}{n}$$

$$s_x = \sqrt{\frac{1}{n-1}\sum (x_i - \bar{x})^2}$$

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1) + (n_2 - 1)}}$$

$$\hat{y} = b_o + b_1 x$$

$$b_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

$$b_o = \bar{y} - b_1 \bar{x}$$

$$r = \frac{1}{n-1}\sum \left(\frac{x_i - \bar{x}}{s_x}\right)\left(\frac{y_i - \bar{y}}{s_y}\right)$$

$$b_1 = r\frac{s_y}{s_x}$$

$$s_{b_1} = \frac{\sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n-2}}}{\sqrt{\sum (x_i - \bar{x})^2}}$$

(II) Probability

P(A $\bigcup$ B) = P(A) + P(B) – P(A ∩ B)

P(A | B) = $\dfrac{P(A \bigcap B)}{P(B)}$

E(X) = $\mu_x = \sum x_i p_i$

Var(X) = $\sigma_x^2 = \sum (x_i - \mu_x)^2 p_i$

If X has a binomial distribution with Parameters *n* and *p*, then:

P(X = *k*) = $\binom{n}{k} p^k (1-p)^{n-k}$

$\mu_x = np$

$$\sigma_x = \sqrt{np(1-p)}$$

$$\mu_{\hat{p}} = p$$

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

If $\bar{x}$ is the mean of a random sample of size *n* from an infinite population with mean μ and standard deviation σ, then:

$$\mu_{\bar{x}} = \mu$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

(III) Inferential Statistics

Standardized test statistic: $\dfrac{\text{statistic} - \text{parameter}}{\text{standard deviation of statistic}}$

Confidence interval: statistic ± (critical value) • (standard deviation of statistic)

Single–Sample

Statistic	Standard Deviation Of Statistic
Sample Mean	$\dfrac{\sigma}{\sqrt{n}}$
Sample Proportion	$\sqrt{\dfrac{p(1-p)}{n}}$

Two–Sample

Statistic	Standard Deviation Of Statistic
Difference of sample means	$\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}$ Special case when $\sigma_1 = \sigma_2$ $\sigma\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}$
Difference of sample proportions	$\sqrt{\dfrac{p_1(1-p_1)}{n_1}+\dfrac{p_2(1-p_2)}{n_2}}$ Special case when $p_1 = p_2$ $\sqrt{p(1-p)}\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}$

Chi-square test statistic = $\sum \dfrac{(observed - \exp ected)^2}{\exp ected}$