TABLES AND FORMULAS FOR MOORE
Basic Practice of Statistics

Exploring Data: Distributions

- Look for overall pattern (shape, center, spread) and deviations (outliers).
- Mean (use a calculator):

$$\overline{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{1}{n}\sum x_i$$

- Standard deviation (use a calculator):

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

- Median: Arrange all observations from smallest to largest. The median M is located $(n+1)/2$ observations from the beginning of this list.
- Quartiles: The first quartile Q_1 is the median of the observations whose position in the ordered list is to the left of the location of the overall median. The third quartile Q_3 is the median of the observations to the right of the location of the overall median.
- Five-number summary:

$$\text{Minimum}, \quad Q_1, \quad M, \quad Q_3, \quad \text{Maximum}$$

- Standardized value of x:

$$z = \frac{x - \mu}{\sigma} \qquad \texttt{[population]}$$

$$z = \frac{x - \overline{x}}{s} \qquad \texttt{[sample]}$$

Exploring Data: Relationships

- Look for overall pattern (form, direction, strength) and deviations (outliers, influential observations).

Un-standardize: $x = \mu + z_p \cdot \sigma$

- Correlation (use a calculator):

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

- Least-squares regression line (use a calculator): $\hat{y} = a + bx$ with slope $b = rs_y/s_x$ and intercept $a = \overline{y} - b\overline{x}$
- Residuals:

$$\text{residual} = \text{observed } y - \text{predicted } y = y - \hat{y}$$

Producing Data

- Simple random sample: Choose an SRS by giving every individual in the population a numerical label and using Table B of random digits to choose the sample.
- Randomized comparative experiments:

$$\begin{tikzcd}
 & \text{Group 1} \arrow[r] & \text{Treatment 1} \arrow[dr] & \\
\text{Random Allocation} \arrow[ur] \arrow[dr] & & & \text{Observe Response} \\
 & \text{Group 2} \arrow[r] & \text{Treatment 2} \arrow[ur] &
\end{tikzcd}$$

Probability and Sampling Distributions

- Probability rules:
  - Any probability satisfies $0 \le P(A) \le 1$.
  - The sample space S has probability $P(S) = 1$.
  - If events A and B are disjoint, $P(A \text{ or } B) = P(A) + P(B)$.
  - For any event A, $P(A \text{ does not occur}) = 1 - P(A)$

Marginal proportion: $\hat{p} = \dfrac{\text{marginal total}}{\text{table total}}$

Conditional row proportion: $\hat{p} = \dfrac{\text{cell count}}{\text{row total}}$

Conditional column proportion: $\hat{p} = \dfrac{\text{cell count}}{\text{column total}}$

- Sampling distribution of a sample mean:
  - $\overline{x}$ has mean μ and standard deviation $\sigma/\sqrt{n}$.
  - $\overline{x}$ has a Normal distribution if the population distribution is Normal.
  - Central limit theorem: $\overline{x}$ is approximately Normal when n is large.

Basics of Inference

- z confidence interval for a population mean (σ known, SRS from Normal population):

$$\overline{x} \pm z^* \frac{\sigma}{\sqrt{n}} \qquad z^* \text{ from } N(0,1)$$

- Sample size for desired margin of error m:

$$n = \left(\frac{z^*\sigma}{m}\right)^2$$

- z test statistic for $H_0 : \mu = \mu_0$ (σ known, SRS from Normal population):

$$z = \frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}} \qquad P\text{-values from } N(0,1)$$

Inference About Means

- t confidence interval for a population mean (SRS from Normal population):

$$\overline{x} \pm t^* \frac{s}{\sqrt{n}} \qquad t^* \text{ from } t(n-1)$$

- t test statistic for $H_0 : \mu = \mu_0$ (SRS from Normal population):

$$t = \frac{\overline{x} - \mu_0}{s/\sqrt{n}} \qquad P\text{-values from } t(n-1)$$

- Matched pairs: To compare the responses to the two treatments, apply the one-sample t procedures to the observed differences.

- Two-sample t confidence interval for $\mu_1 - \mu_2$ (independent SRSs from Normal populations):

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

with conservative t^* from t with df the smaller of $n_1 - 1$ and $n_2 - 1$ (or use software).

- Two-sample t test statistic for $H_0 : \mu_1 = \mu_2$ (independent SRSs from Normal populations):

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

with conservative P-values from t with df the smaller of $n_1 - 1$ and $n_2 - 1$ (or use software).

Inference About Proportions

- Sampling distribution of a sample proportion: when the population and the sample size are both large and p is not close to 0 or 1, $\hat{p}$ is approximately Normal with mean p and standard deviation $\sqrt{p(1-p)/n}$.

- Large-sample z confidence interval for p:

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \qquad z^* \text{ from } N(0,1)$$

Plus four to greatly improve accuracy: use the same formula after adding 2 successes and two failures to the data.

- z test statistic for $H_0 : p = p_0$ (large SRS):

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \qquad P\text{-values from } N(0,1)$$

- Sample size for desired margin of error m:

$$n = \left(\frac{z^*}{m}\right)^2 p^*(1-p^*)$$

where p^* is a guessed value for p or $p^* = 0.5$.

- Large-sample z confidence interval for $p_1 - p_2$:

$$(\hat{p}_1 - \hat{p}_2) \pm z^* \text{SE} \qquad z^* \text{ from } N(0,1)$$

where the standard error of $\hat{p}_1 - \hat{p}_2$ is

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

Plus four to greatly improve accuracy: use the same formulas after adding one success and one failure to each sample.

- Two-sample z test statistic for $H_0 : p_1 = p_2$ (large independent SRSs):

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

where $\hat{p}$ is the pooled proportion of successes.

The Chi-Square Test

- Expected count for a cell in a two-way table:

$$\text{expected count} = \frac{\text{row total} \times \text{column total}}{\text{table total}}$$

- Chi-square test statistic for testing whether the row and column variables in an $r \times c$ table are unrelated (expected cell counts not too small):

$$X^2 = \sum \frac{(\text{observed count} - \text{expected count})^2}{\text{expected count}}$$

with P-values from the chi-square distribution with df $= (r-1) \times (c-1)$.

- Describe the relationship using percents, comparison of observed with expected counts, and terms of X^2.

Inference for Regression

- Conditions for regression inference: n observations on x and y. The response y for any fixed x has a Normal distribution with mean given by the true regression line $\mu_y = \alpha + \beta x$ and standard deviation σ. Parameters are α, β, σ.

- Estimate α by the intercept a and β by the slope b of the least-squares line. Estimate σ by the regression standard error:

$$s = \sqrt{\frac{1}{n-2}\sum \text{residual}^2}$$

Use software for all standard errors in regression.

- t confidence interval for regression slope β:

$$b \pm t^* \text{SE}_b \qquad t^* \text{ from } t(n-2)$$

- t test statistic for no linear relationship, $H_0 : \beta = 0$:

$$t = \frac{b}{\text{SE}_b} \qquad P\text{-values from } t(n-2)$$

- t confidence interval for mean response μ_y when $x = x^*$:

$$\hat{y} \pm t^* \text{SE}_{\hat{\mu}} \qquad t^* \text{ from } t(n-2)$$

- t prediction interval for an individual observation y when $x = x^*$:

$$\hat{y} \pm t^* \text{SE}_{\hat{y}} \qquad t^* \text{ from } t(n-2)$$

One-way Analysis of Variance: Comparing Several Means

- ANOVA F tests whether all of I populations have the same mean, based on independent SRSs from I Normal populations with the same σ. P-values come from the F distribution with $I-1$ and $N-I$ degrees of freedom, where N is the total observations in all samples.

- Describe the data using the I sample means and standard deviations and side-by-side graphs of the samples.

- The ANOVA F test statistic (use software) is $F = \text{MSG}/\text{MSE}$, where

$$\begin{aligned} \text{MSG} &= \frac{n_1(\overline{x}_1 - \overline{x})^2 + \cdots + n_I(\overline{x}_I - \overline{x})^2}{I-1} \\ \text{MSE} &= \frac{(n_1-1)s_1^2 + \cdots + (n_I-1)s_I^2}{N-I} \end{aligned}$$

Table entry for z is the area under the standard Normal curve to the left of z.

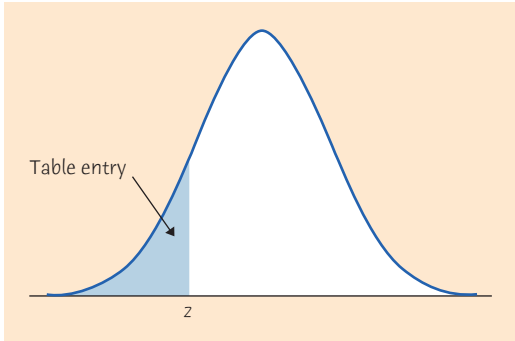


TABLE A Standard Normal cumulative proportions

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−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

Table entry for z is the area under the standard Normal curve to the left of z.

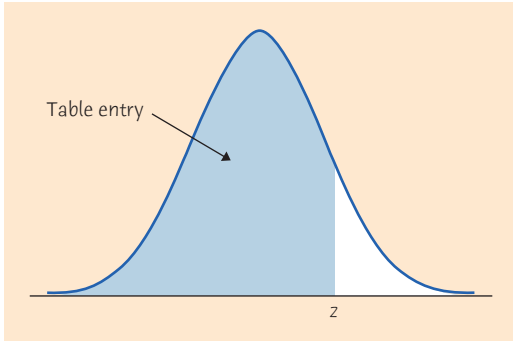


TABLE A Standard Normal cumulative proportions (*continued*)

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

Table entry for C is the critical value t^* required for confidence level C. To approximate one- and two-sided P-values, compare the value of the t statistic with the critical values of t^* that match the P-values given at the bottom of the table.

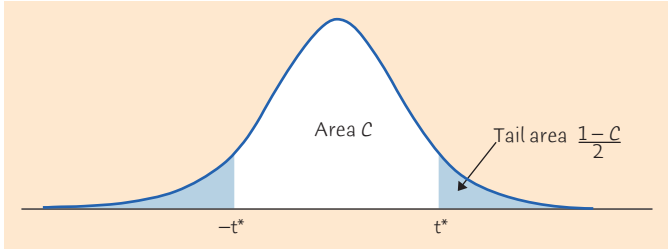
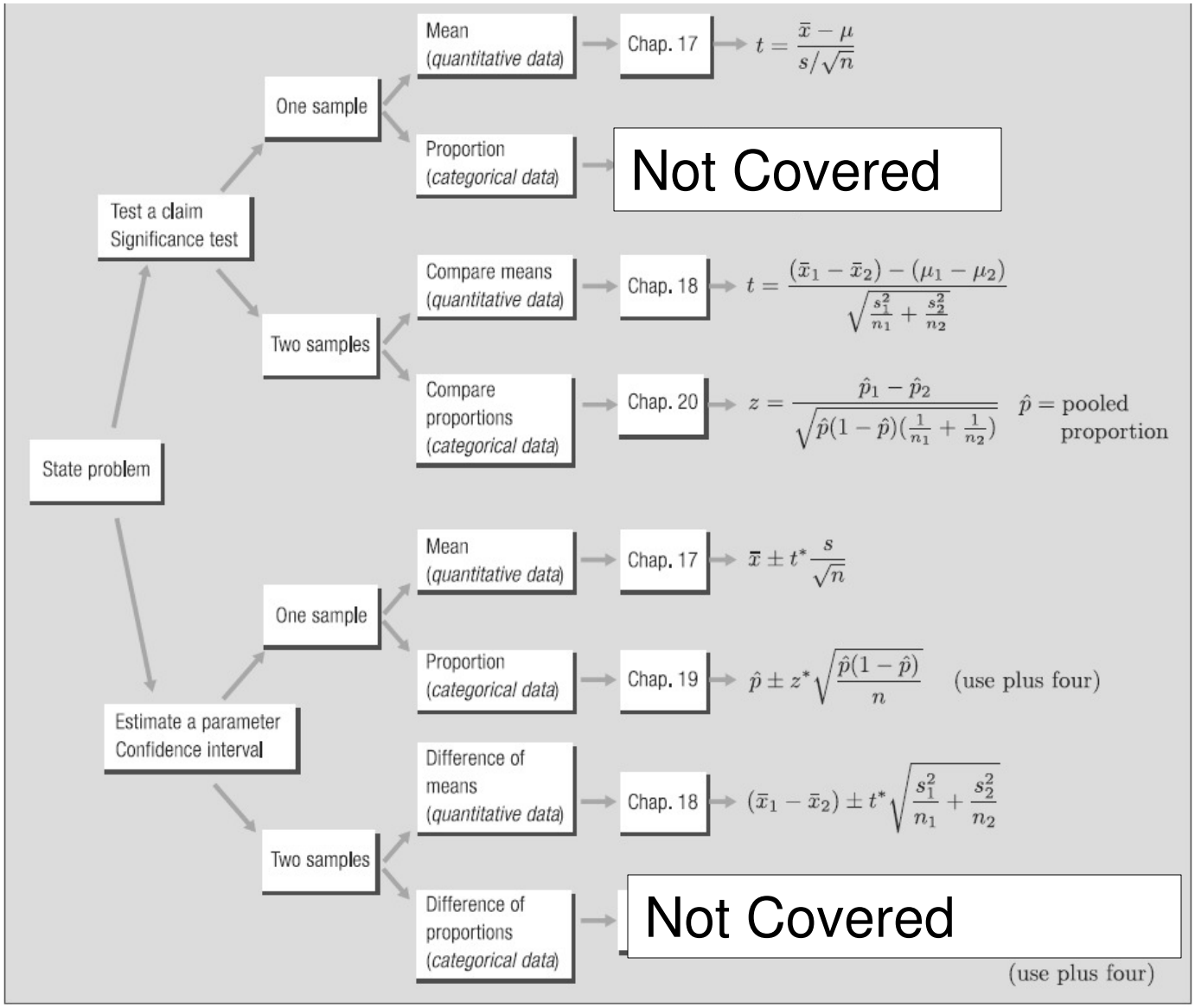


TABLE C *t* distribution critical values

DEGREES OF FREEDOM	CONFIDENCE LEVEL *C*											
	50%	60%	70%	80%	90%	95%	96%	98%	99%	99.5%	99.8%	99.9%
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	127.3	318.3	636.6
2	0.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	0.765	0.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878	3.197	3.611	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861	3.174	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845	3.153	3.552	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831	3.135	3.527	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819	3.119	3.505	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807	3.104	3.485	3.768
24	0.685	0.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797	3.091	3.467	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787	3.078	3.450	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.162	2.479	2.779	3.067	3.435	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.158	2.473	2.771	3.057	3.421	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.154	2.467	2.763	3.047	3.408	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.150	2.462	2.756	3.038	3.396	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.147	2.457	2.750	3.030	3.385	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.123	2.423	2.704	2.971	3.307	3.551
50	0.679	0.849	1.047	1.299	1.676	2.009	2.109	2.403	2.678	2.937	3.261	3.496
60	0.679	0.848	1.045	1.296	1.671	2.000	2.099	2.390	2.660	2.915	3.232	3.460
80	0.678	0.846	1.043	1.292	1.664	1.990	2.088	2.374	2.639	2.887	3.195	3.416
100	0.677	0.845	1.042	1.290	1.660	1.984	2.081	2.364	2.626	2.871	3.174	3.390
1000	0.675	0.842	1.037	1.282	1.646	1.962	2.056	2.330	2.581	2.813	3.098	3.300
z^*	0.674	0.841	1.036	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.091	3.291
One-sided P	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
Two-sided P	.50	.40	.30	.20	.10	.05	.04	.02	.01	.005	.002	.001


State problem
Test a claim
Significance test
One sample
Mean
(quantitative data)
Chap. 17
$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$
Proportion
(categorical data)
Not Covered
Two samples
Compare means
(quantitative data)
Chap. 18
$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$
Compare
proportions
(categorical data)
Chap. 20
$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}}$
$\hat{p}$ = pooled proportion
Estimate a parameter
Confidence interval
One sample
Mean
(quantitative data)
Chap. 17
$\bar{x} \pm t^* \frac{s}{\sqrt{n}}$
Proportion
(categorical data)
Chap. 19
$\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$
(use plus four)
Two samples
Difference of
means
(quantitative data)
Chap. 18
$(\bar{x}_1 - \bar{x}_2) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$
Difference of
proportions
(categorical data)
Not Covered
(use plus four)