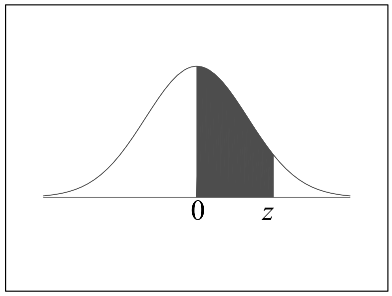
COMM 215 Final Review

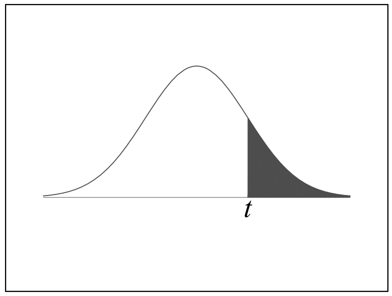
Standard Normal Distribution Table



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990
3.1	.4990	.4991	.4991	.4991	.4992	.4992	.4992	.4992	.4993	.4993
3.2	.4993	.4993	.4994	.4994	.4994	.4994	.4994	.4995	.4995	.4995
3.3	.4995	.4995	.4995	.4996	.4996	.4996	.4996	.4996	.4996	.4997
3.4	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4998
3.5	.4998	.4998	.4998	.4998	.4998	.4998	.4998	.4998	.4998	.4998

Gilles Cazelais. Typeset with LaTeX on April 20, 2006.

t-Distribution Table



The shaded area is equal to α for $t = t_\alpha$.

df	$t_{.100}$	$t_{.050}$	$t_{.025}$	$t_{.010}$	$t_{.005}$
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
32	1.309	1.694	2.037	2.449	2.738
34	1.307	1.691	2.032	2.441	2.728
36	1.306	1.688	2.028	2.434	2.719
38	1.304	1.686	2.024	2.429	2.712
∞	1.282	1.645	1.960	2.326	2.576

Sampling Distributions

Consider a population of $N = 8$ numbers: $3, 6, 9, 12, 15, 18, 21, 24$. The mean and standard deviation of the population are

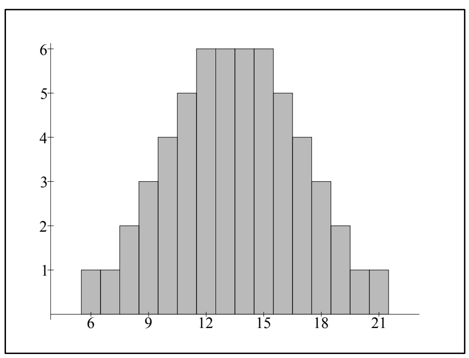
$$\mu = 13.5 \quad \text{and} \quad \sigma = 6.8739.$$

There are $C(8, 3) = 56$ possible samples of size $n = 3$. All these samples and their sample mean $\bar{x}$ are listed in the following table.

Sample Values	$\bar{x}$	Sample Values	$\bar{x}$	Sample Values	$\bar{x}$	Sample Values	$\bar{x}$
3, 6, 9	6	3, 12, 24	13	6, 12, 21	13	9, 15, 24	16
3, 6, 12	7	3, 15, 18	12	6, 12, 24	14	9, 18, 21	16
3, 6, 15	8	3, 15, 21	13	6, 15, 18	13	9, 18, 24	17
3, 6, 18	9	3, 15, 24	14	6, 15, 21	14	9, 21, 24	18
3, 6, 21	10	3, 18, 21	14	6, 15, 24	15	12, 15, 18	15
3, 6, 24	11	3, 18, 24	15	6, 18, 21	15	12, 15, 21	16
3, 9, 12	8	3, 21, 24	16	6, 18, 24	16	12, 15, 24	17
3, 9, 15	9	6, 9, 12	9	6, 21, 24	17	12, 18, 21	17
3, 9, 18	10	6, 9, 15	10	9, 12, 15	12	12, 18, 24	18
3, 9, 21	11	6, 9, 18	11	9, 12, 18	13	12, 21, 24	19
3, 9, 24	12	6, 9, 21	12	9, 12, 21	14	15, 18, 21	18
3, 12, 15	10	6, 9, 24	13	9, 12, 24	15	15, 18, 24	19
3, 12, 18	11	6, 12, 15	11	9, 15, 18	14	15, 21, 24	20
3, 12, 21	12	6, 12, 18	12	9, 15, 21	15	18, 21, 24	21

These values of $\bar{x}$ have the following distribution.

$\bar{x}$	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
Frequency	1	1	2	3	4	5	6	6	6	6	5	4	3	2	1	1



If we now think of the sample mean $\bar{x}$ as a random variable, its mean is $\mu_{\bar{x}} = 13.5$, its standard deviation is $\sigma_{\bar{x}} = 3.3541$, and its distribution looks roughly like a normal. As the sample size n gets larger, the sample mean gets closer to a normal distribution. This follows from the following fundamental theorem.

Central Limit Theorem. *If random samples of size n are drawn from any population with finite mean μ and standard deviation σ, then for large n, the sampling distribution of the sample mean $\bar{x}$ is approximately normally distributed with*

$$\mu_{\bar{x}} = \mu \quad \textit{and} \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}.$$

The approximation gets more accurate as n increases.

How large should n be to get a good approximation? Usually $n = 30$ is enough to get a good normal approximation. If the sampled population is normal, then $\bar{x}$ is always normal, no matter what the sample size n is. In general $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} \approx \frac{\sigma}{\sqrt{n}}$ when N is much larger than n.

Chi-Square Distribution

Inference about a Population Variance

If the population from which the sample is selected is (approximately) normally distributed, then

$$\frac{(n-1)s^2}{\sigma^2}$$

has a chi-square distribution with $n-1$ degrees of freedom.

A $(1-\alpha)100\%$ Confidence Interval for σ^2 is of the form

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{(1-\alpha/2)}}.$$

The confidence interval for σ can be obtained by taking the square root of the two limits of the above interval.

Test about a Population Variance

Null Hypothesis: $H_0\colon \sigma^2 = \sigma_0^2$

Test Statistic: $\chi^2 = \dfrac{(n-1)s^2}{\sigma_0^2}$

Alternative Hypothesis	Rejection region at the significance level α
$H_a\colon \sigma^2 > \sigma_0^2$	$\chi^2 > \chi^2_\alpha$ (upper-tailed test)
$H_a\colon \sigma^2 < \sigma_0^2$	$\chi^2 < \chi^2_\alpha$ (lower-tailed test)
$H_a\colon \sigma^2 \neq \sigma_0^2$	$\chi^2 > \chi^2_{\alpha/2}$ or $\chi^2 < \chi^2_{(1-\alpha/2)}$ (two-tailed test)

Goodness-of-fit Test

A **Multinomial Experiment** is an experiment with the following properties.

1. It consists of n identical independent trials.
2. Each trial results in one of k possible outcomes.
3. The probabilities p_i of the possible outcomes remain constant for each trial.

The test statistic for a goodness-of-fit test of a multinomial experiment is

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

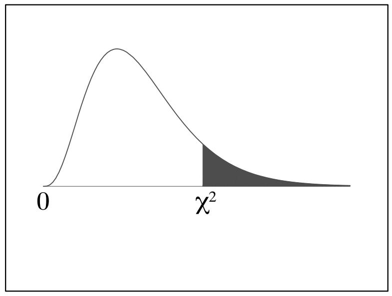
where:

$O_i =$ observed frequency of category i, and $E_i =$ expected frequency of category $i = np_i$.

In a goodness-of-fit test, the degrees of freedom are $df = k-1$. It is always an upper-tailed test. The number of trials n should be large enough so that $np_i > 5$ for all i.

The p-value is $P(\chi^2 > \chi^2_{\text{obs}})$. It can be calculated using the HP as follows: df, χ^2_{obs}, UTPC.

Chi-Square Distribution Table



The shaded area is equal to α for $\chi^2 = \chi^2_\alpha$.

df	$\chi^2_{.995}$	$\chi^2_{.990}$	$\chi^2_{.975}$	$\chi^2_{.950}$	$\chi^2_{.900}$	$\chi^2_{.100}$	$\chi^2_{.050}$	$\chi^2_{.025}$	$\chi^2_{.010}$	$\chi^2_{.005}$
1	0.000	0.000	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	11.070	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	28.300
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	14.041	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	43.195	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	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.336
30	13.787	14.953	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

Appendix K

The F Distribution

The F distribution is an asymmetric distribution that has a minimum value of 0, but no maximum value. The curve reaches a peak not far to the right of 0, and then gradually approaches the horizontal axis the larger the F value is. The F distribution approaches, but never quite touches the horizontal axis.

The F distribution has two degrees of freedom, d_1 for the numerator, d_2 for the denominator. For each combination of these degrees of freedom there is a different F distribution. The F distribution is most spread out when the degrees of freedom are small. As the degrees of freedom increase, the F distribution the F distribution is less dispersed.

Figure 1.1 shows the shape of the distribution. The F value is on the horizontal axis, with the probability for each F value being represented by the vertical axis. The shaded area in the diagram represents the level of significance α shown in the table.

There is a different F distribution for each combination of the degrees of freedom of the numerator and denominator. Since there are so many F distributions, the F tables are organized somewhat differently than the tables for the other distributions. The three tables which follow are organized by the level of significance. The first table gives F values for that are associated with $\alpha = 0.10$ of the area in the right tail of the distribution. The second table gives the F values for $\alpha = 0.05$ of the area in the right tail, and the third table gives F values for the $\alpha = 0.01$ level of significance. In each of these tables, the F values are given for various combinations of degrees of freedom.

In order to use the F table, first select the significance level to be used,

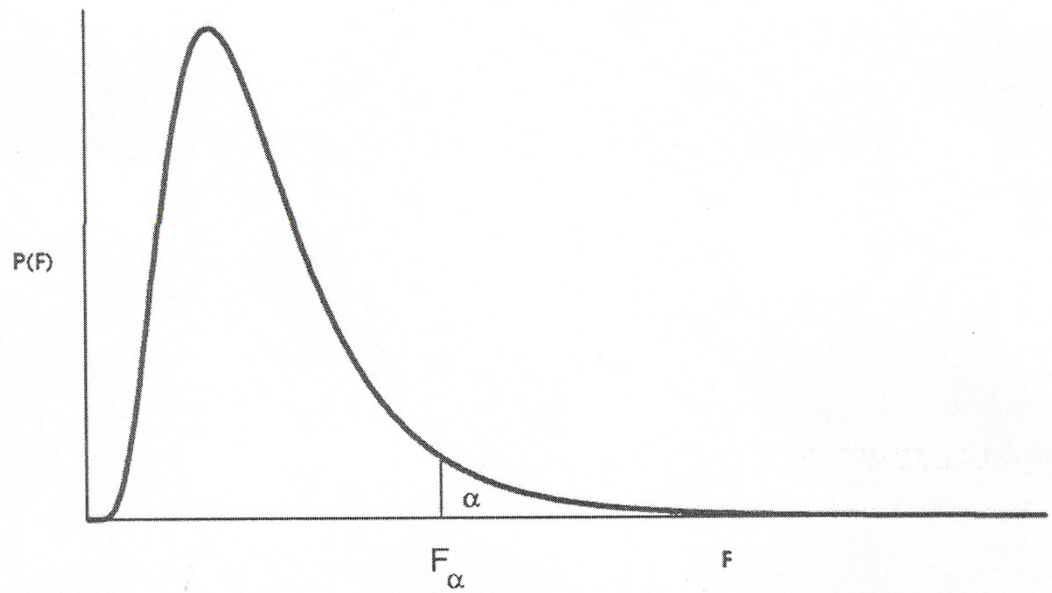


Figure K.1: The F distribution

F distribution critical value landmarks

Table entries are critical values for F^* with probably p in right tail of the distribution.

Figure of *F* distribution (like in Moore, 2004, p. 656) here.

Degrees of freedom in denominator (df2)	*p*	Degrees of freedom in numerator (df1) 1	2	3	4	5	6	7	8	12	24	1000
1	0.100	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	60.71	62.00	63.30
	0.050	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	243.9	249.1	254.2
	0.025	647.8	799.5	864.2	899.6	921.8	937.1	948.2	956.6	976.7	997.3	1017.8
	0.010	4052	4999	5404	5624	5764	5859	5928	5981	6107	6234	6363
	0.001	405312	499725	540257	562668	576496	586033	593185	597954	610352	623703	636101
2	0.100	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.41	9.45	9.49
	0.050	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.41	19.45	19.49
	0.025	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.41	39.46	39.50
	0.010	98.50	99.00	99.16	99.25	99.30	99.33	99.36	99.38	99.42	99.46	99.50
	0.001	998.38	998.84	999.31	999.31	999.31	999.31	999.31	999.31	999.31	999.31	999.31
3	0.100	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.22	5.18	5.13
	0.050	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.74	8.64	8.53
	0.025	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.34	14.12	13.91
	0.010	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.05	26.60	26.14
	0.001	167.06	148.49	141.10	137.08	134.58	132.83	131.61	130.62	128.32	125.93	123.52
4	0.100	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.90	3.83	3.76
	0.050	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	5.91	5.77	5.63
	0.025	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.75	8.51	8.26
	0.010	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.37	13.93	13.47
	0.001	74.13	61.25	56.17	53.43	51.72	50.52	49.65	49.00	47.41	45.77	44.09
5	0.100	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.27	3.19	3.11
	0.050	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.68	4.53	4.37
	0.025	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.52	6.28	6.02
	0.010	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	9.89	9.47	9.03
	0.001	47.18	37.12	33.20	31.08	29.75	28.83	28.17	27.65	26.42	25.13	23.82
6	0.100	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.90	2.82	2.72
	0.050	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.00	3.84	3.67
	0.025	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.37	5.12	4.86
	0.010	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.72	7.31	6.89
	0.001	35.51	27.00	23.71	21.92	20.80	20.03	19.46	19.03	17.99	16.90	15.77
7	0.100	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.67	2.58	2.47
	0.050	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.57	3.41	3.23
	0.025	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.67	4.41	4.15
	0.010	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.47	6.07	5.66
	0.001	29.25	21.69	18.77	17.20	16.21	15.52	15.02	14.63	13.71	12.73	11.72
8	0.100	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.50	2.40	2.30
	0.050	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.28	3.12	2.93
	0.025	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.20	3.95	3.68
	0.010	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.67	5.28	4.87
	0.001	25.41	18.49	15.83	14.39	13.48	12.86	12.40	12.05	11.19	10.30	9.36
9	0.100	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.38	2.28	2.16
	0.050	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.07	2.90	2.71
	0.025	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	3.87	3.61	3.34
	0.010	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.11	4.73	4.32
	0.001	22.86	16.39	13.90	12.56	11.71	11.13	10.70	10.37	9.57	8.72	7.84

Critical values computed with Excel 9.0

Degrees of freedom in denominator (df2)	*p*	Degrees of freedom in numerator (df1)										
	p	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**12**	**24**	**1000**
10	0.100	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.28	2.18	2.06
	0.050	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	2.91	2.74	2.54
	0.025	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.62	3.37	3.09
	0.010	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.71	4.33	3.92
	0.001	21.04	14.90	12.55	11.28	10.48	9.93	9.52	9.20	8.45	7.64	6.78
12	0.100	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.15	2.04	1.91
	0.050	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.69	2.51	2.30
	0.025	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.28	3.02	2.73
	0.010	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.16	3.78	3.37
	0.001	18.64	12.97	10.80	9.63	8.89	8.38	8.00	7.71	7.00	6.25	5.44
14	0.100	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.05	1.94	1.80
	0.050	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.53	2.35	2.14
	0.025	6.30	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.05	2.79	2.50
	0.010	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	3.80	3.43	3.02
	0.001	17.14	11.78	9.73	8.62	7.92	7.44	7.08	6.80	6.13	5.41	4.62
16	0.100	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	1.99	1.87	1.72
	0.050	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.42	2.24	2.02
	0.025	6.12	4.69	4.08	3.73	3.50	3.34	3.22	3.12	2.89	2.63	2.32
	0.010	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.55	3.18	2.76
	0.001	16.12	10.97	9.01	7.94	7.27	6.80	6.46	6.20	5.55	4.85	4.08
18	0.100	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	1.93	1.81	1.66
	0.050	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.34	2.15	1.92
	0.025	5.98	4.56	3.95	3.61	3.38	3.22	3.10	3.01	2.77	2.50	2.20
	0.010	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.37	3.00	2.58
	0.001	15.38	10.39	8.49	7.46	6.81	6.35	6.02	5.76	5.13	4.45	3.69
20	0.100	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.89	1.77	1.61
	0.050	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.28	2.08	1.85
	0.025	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.68	2.41	2.09
	0.010	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.23	2.86	2.43
	0.001	14.82	9.95	8.10	7.10	6.46	6.02	5.69	5.44	4.82	4.15	3.40
30	0.100	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.77	1.64	1.46
	0.050	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.09	1.89	1.63
	0.025	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.41	2.14	1.80
	0.010	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	2.84	2.47	2.02
	0.001	13.29	8.77	7.05	6.12	5.53	5.12	4.82	4.58	4.00	3.36	2.61
50	0.100	2.81	2.41	2.20	2.06	1.97	1.90	1.84	1.80	1.68	1.54	1.33
	0.050	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	1.95	1.74	1.45
	0.025	5.34	3.97	3.39	3.05	2.83	2.67	2.55	2.46	2.22	1.93	1.56
	0.010	7.17	5.06	4.20	3.72	3.41	3.19	3.02	2.89	2.56	2.18	1.70
	0.001	12.22	7.96	6.34	5.46	4.90	4.51	4.22	4.00	3.44	2.82	2.05
100	0.100	2.76	2.36	2.14	2.00	1.91	1.83	1.78	1.73	1.61	1.46	1.22
	0.050	3.94	3.09	2.70	2.46	2.31	2.19	2.10	2.03	1.85	1.63	1.30
	0.025	5.18	3.83	3.25	2.92	2.70	2.54	2.42	2.32	2.08	1.78	1.36
	0.010	6.90	4.82	3.98	3.51	3.21	2.99	2.82	2.69	2.37	1.98	1.45
	0.001	11.50	7.41	5.86	5.02	4.48	4.11	3.83	3.61	3.07	2.46	1.64
1000	0.100	2.71	2.31	2.09	1.95	1.85	1.78	1.72	1.68	1.55	1.39	1.08
	0.050	3.85	3.00	2.61	2.38	2.22	2.11	2.02	1.95	1.76	1.53	1.11
	0.025	5.04	3.70	3.13	2.80	2.58	2.42	2.30	2.20	1.96	1.65	1.13
	0.010	6.66	4.63	3.80	3.34	3.04	2.82	2.66	2.53	2.20	1.81	1.16
	0.001	10.89	6.96	5.46	4.65	4.14	3.78	3.51	3.30	2.77	2.16	1.22

Use StaTable, WinPepi > WhatIs, or other reliable software to determine specific *p* values

Hypothesis Testing Notes

Test of **Mean**

1. **State the Hypothesis**
   a. Ho: μ = or ≤ or ≥ depending on the wording
   b. Ha: μ ≠ or < or >
2. **Determine the Critical Value**
   a. use **Z** if given **σ** (population standard deviation)
      - Z_α or $Z_{\frac{\alpha}{2}}$
   b. use **t** if given **s** (sample standard deviation)
      - $t_{\alpha,n-1}$ or $t_{\frac{\alpha}{2},n-1}$
3. **Calculate the test statistic (T-stat)**
   a. $Z^* = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$
   b. $t^* = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$
4. **Conclude**
   a. if |Z*| > |Z|, reject Ho
   b. if |t*| < |t|, do not reject Ho
5. **Calculate the P value**

Test of **Proportion**

1. **State the Hypothesis**
   a. Ho: π = or ≤ or ≥ depending on the wording
   b. Ha: π ≠ or < or >
2. **Determine the Critical Value**
   a. use Z
      - Z_α or $Z_{\frac{\alpha}{2}}$
3. **Calculate the test statistic (T-stat)**
   a. $Z^* = \frac{p-\pi}{\sqrt{\frac{\pi(1-\pi)}{n}}}$
4. **Conclude**
   a. if |Z*| > |Z|, reject Ho
   b. if |Z*| < |Z|, do not reject Ho
5. **Calculate the P value**

Test of **Independence**

1. **State the Hypothesis**
   a. Category x & y are independent
   b. Category x & y are not independent
2. **Determine the Critical Value**
   a. use χ^2
      - ${\chi^2}_{\alpha,(r-1)(c-1)}$
3. **Calculate the test statistic (T-stat)**
   a. $\chi^{2*} = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(o_{ij}-e_{ij})^2}{e_{ij}}$
4. **Conclude**
   a. if $\chi^{2*} > \chi^2$, reject Ho
   b. if $\chi^{2*} < \chi^2$, do not reject Ho
5. **Calculate the P value**

Test of **Goodness of Fit**

1. **State the Hypothesis**
   a. Series follow some state distribution
   b. Series follow some other distribution
2. **Determine the Critical Value**
   a. use χ^2
      i. ${\chi^2}_{\alpha,k-1}$
3. **Calculate the test statistic (T-stat)**
   a. $\chi^{2*} = \sum_{i=1}^{k}\frac{(o_i-e_i)^2}{e_i}$
4. **Conclude**
   a. if $\chi^{2*} > \chi^2$, reject Ho
   b. if $\chi^{2*} < \chi^2$, do not reject Ho
5. **Calculate the P value**

Test of **Rho**

1. **State the Hypothesis**
    a. Ho: $\rho = 0$
    b. Ha: $\rho \neq 0$
2. **Determine the Critical Value**
    a. use t
        i. $t_{\frac{\alpha}{2},n-2}$
3. **Calculate the test statistic (T-stat)**
    a. $t^* = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}}$
4. **Conclude**
    a. if $|t^*| > |t|$, reject Ho
    b. if $|t^*| < |t|$, do not reject Ho
5. **Calculate the P value**

Test of **Rho²**

1. **State the Hypothesis**
    a. Ho: $\rho^2 = 0$
    b. Ha: $\rho^2 > 0$
2. **Determine the Critical Value**
    a. use F
        i. $F_{\alpha,d_1/d_2}$
3. **Calculate the test statistic (T-stat)**
    a. $F^* = \frac{\left(\frac{SSR}{k}\right)}{\left(\frac{SSE}{n-k-1}\right)} = \frac{MSR}{MSE}$
4. **Conclude**
    a. if $|F^*| > |F|$, reject Ho
    b. if $|F^*| < |F|$, do not reject Ho
5. **Calculate the P value**
    a. Found in printout

Test of **Slope**

1. **State the Hypothesis**
    a. Ho: β_i = or ≤ or ≥ depending on the wording
    b. Ha: β_i ≠ or < or >
2. **Determine the Critical Value**
    a. use t
        - $t_{\alpha,n-k-1}$ or $t_{\frac{\alpha}{2},n-k-1}$
3. **Calculate the test statistic (T-stat)**
    a. $t^* = \frac{b_1-\beta_1}{s_{b_i}}$
4. **Conclude**
    a. if $|t^*| > |t|$, reject Ho
    b. if $|t^*| < |t|$, do not reject Ho
5. **Calculate P Value**
    a. Found in printout

Test of **Overall Model**

1. **State the Hypothesis**
    a. Ho: $\beta_1 = \beta_2 = \beta_3 = \beta_i = \ldots = 0$
    b. Ha: at least one $\beta_i \neq$
2. **Determine the Critical Value**
    a. use F
        i. $F_{\alpha,d_1/d_2}$
3. **Calculate the test statistic (T-stat)**
    a. $F^* = \frac{\left(\frac{SSR}{k}\right)}{\left(\frac{SSE}{n-k-1}\right)} = \frac{MSR}{MSE}$
4. **Conclude**
    a. if $|F^*| > |F|$, reject Ho
    b. if $|F^*| < |F|$, do not reject Ho
5. **Calculate the P value**
    a. Found in printout

Regression Notes

- $r_{xy} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} = \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{\sqrt{\left[\sum x_i{}^2 - \frac{\left(\sum x_i\right)^2}{n}\right]\left[\sum y_i{}^2 - \frac{\left(\sum y_i\right)^2}{n}\right]}} = \frac{SS_{xy}}{SS_{xx} SS_{yy}}$

- $\hat{y} = b_0 + b_1 x$

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

- $b_0 = \bar{y} - b_1 \bar{x}$

- $r^2 = \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2} = \frac{SSR}{SST} = \left(1 - \frac{SSE}{SST}\right) = \left(1 - \frac{\sum y_i{}^2 - b_0 \sum y_i - b_1 \sum x_i y_i}{\sum y_i{}^2 - \frac{\left(\sum y_i\right)^2}{n}}\right)$

- $s_{b_1} = \sqrt{\frac{\frac{\sum(y_i - \hat{y}_i)^2}{n-k-1}}{\sum(x_i - \bar{x})^2}} = \sqrt{\frac{\frac{\sum(y_i - \hat{y}_i)^2}{n-k-1}}{\sum x_i^2 - \frac{(\sum x)^2}{n}}} = \sqrt{\frac{\frac{SSE}{n-k-1}}{SS_{xx}}} = \frac{s_\varepsilon}{\sqrt{SS_{xx}}}$

Regression Statistics	
Multiple R	R
R Square	R^2
Adjusted R Square	$1-(1-R^2)(n-1/n-k-1)$
Standard Error	s_e
Observations	n

ANOVA

source	df	SS	MS	F	Significance
Regression	k	SSR	MSR = SSR/k	MSR/MSE	p-value
Residual (ε)	n-k-1	SSE	MSE = SSE/n-k-1		
Total	n-1	SST			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	b_0	s_{b0}	$(b_0-B_0)/s_{b0}$	-	$b_0 \pm t_{\alpha/2} s_{b0}$	
X_1	b_1	s_{b1}	$(b_0-B_1)/s_{b1}$	-	$b_1 \pm t_{\alpha/2} s_{b1}$	
X_2	b_2	s_{b2}	$(b_0-B_2)/s_{b2}$	-	$b_2 \pm t_{\alpha/2} s_{b2}$	
X_3	b_3	s_{b3}	$(b_0-B_3)/s_{b3}$	-	$b_3 \pm t_{\alpha/2} s_{b3}$	

1. A recent survey on Internet usage included 400 respondents. The results suggested that average time spent by the respondents on Internet is 3.6 hours per day. Standard deviation based on the sample was 1.6 hours.
    a. Construct 95% confidence interval for the average daily amount of time people spend on the Internet.

    b. Assume that the population mean and standard deviation are actually 3.6 hours and 1.6 hours respectively, and that you would like to half the margin of error you obtained in part a. How many more responses you would have to collect? Use a 95% confidence interval.

c. Assume 348 respondents indicated that they use Internet at least 2 hours each day. Provide an interval estimate for the proportion of people using the Internet for 2 hours or more per day. Use 90% confidence level.

d. The Internet provider claims that at least 95% of the population uses the internet for at least 2 hours per day. Based on the given data what can you say about the Internet provider's claim? Justify your answer using the alpha = 5%.

2. As Internet usage proliferates, so do questions of security and confidentiality of personal information, including such things as social security and credit card numbers. Acme security agency surveyed 1000 Canadian adults and asked them under what circumstances they would give personal information to a company. Twenty-nine percent (29%) said they would never give personal data to a company, while 51% said they would if the company had strict privacy guidelines in place.

    a. Verify that the sample size is large enough to construct a valid confidence interval for p, the proportion of all Canadian adults who would never give personal information to a company.

    b. Construct a 95% confidence interval for p and interpret your result in the context of the problem.

3. A new process for producing synthetic diamonds can be operated at a profitable level only if the average weight of the diamonds is greater than 0.5 karat. To evaluate the profitability of the process, six diamonds are generated, with record weights 0.46, 0.61, 0.52, 0.48, 0.57 and 0.54 karat.
   a. Do the six measurements present sufficient evidence to indicate that the average weight of diamonds produced by the process is profitable? At alpha 0.05, explain your conclusion in the context of the problem.

   b. Calculate the p- value and compare your decision with the one in the previous part.

   c. Construct a 95% confidence interval for the true mean weight of the diamonds. Does this bound confirm your conclusion on the profitability of this operation? Explain why.

4. Where people turn to for news differs according to various age groups. A recent study identified from which medium two specific age groups primarily obtain their news. The results of the sample are shown below.

Age Group	**Medium**		
	Television	News Paper	Internet
< 50 years old	400	170	180
≥ 50 years old	260	110	80

a. At the 5 percent level of significance, is there evidence of a relationship between age group and medium where the news are primarily obtained?

b. Which group contributes the least to the decision taken in part a?

c. Construct a 98% confidence interval for the overall proportion of people who primarily obtain their news from the Internet.

5. The "beta coefficient" of a stock is a measure of the stock's volatility (or risk) relative to the market as a whole. Stock with beta coefficients greater than 1 generally bear greater risk (more volatility) than the market, whereas stocks with beta coefficients less than 1 are less risky (less volatile) than the overall market. A random sample of 15 high-technology stocks was selected at the end of 2007, and the mean and standard deviation of the beta coefficients were calculated: sample mean, $X = 1.23$, sample standard deviation, $S=.37$
   a. Set up the appropriate null and alternative hypotheses to test whether the average high-technology stock is riskier than the market as a whole.

   b. Establish the appropriate test statistic and rejection region for the test. Use the probability of making a Type I error to be 5%.

   c. What assumptions are necessary to ensure the validity of the test?

d. Calculate the test statistic and state your conclusion.

e. What is the appropriate p-value associated with this test? Interpret the result.

6. Firms planning to build new plants or make additions to existing facilities have become very conscious of the energy efficiency of proposed new structures and are interested in the relation between yearly energy consumption and the number of square feet of building shell. The accompanying table lists the energy consumption in BTU for 22 buildings that were all subjected to the same climatic conditions. BTU and Shell Area Data is given below:

BTU	3870	1371	2422	672.2	233.1	218.9	354	3135	1470	1408	2201
Shell Area	30001	13530	26060	6355	4576	24680	2621	23350	18770	12220	25490
BTU	2680	337.5	567.5	555.3	239.4	2629	1102	423.5	423.5	1691	1870
Shell Area	23680	5650	8001	6147	2660	19240	10700	9125	6510	13530	18860

Assuming that a linear regression model is appropriate, the following statistics were obtained with the Excel software.

Regression Statistics	
Multiple R	0.82
R Square	0.68
Adjusted R Square	0.66
Standard Error	628.18
Observations	22

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	16,584,975	16,584,975	42.03	0.00
Residual	20	7,892,320	394,616		
Total	21	24,477,296			

	Coefficients	*Standard Error*	*t-Stat*	*P-Value*	*lower 95%*	*lower 95%*
Intercept	-99.05	261.62	-0.38	0.71	-645	447
X	0.10	0.016	6.48	0.00	0.07	0.13

a. Is yearly energy consumption positively related to the shell area of the building? Test at 5% significance level.

b. Find the observed significance level (p-value) of the test of part a. interpret its value.

c. Find the coefficient of determination and interpret the value in the context of this problem.

d. A company wishes to build a new warehouse that will contain 8000 square feet of shell area. Find the predicted value of energy consumption and associated 95% prediction interval. It is given that $\sum x = 311{,}756$ and $\sum x^2 = 5{,}986{,}762{,}178$. Comment on the usefulness of this interval.

e. The application of the present simple linear regression model to the warehouse problem of part d) is appropriate only if certain assumptions can be made about the new warehouse. What are these assumptions?

7. An actuary wanted to develop a model to predict how long individuals will live. After consulting a number of physicians, she collected data on the age at death (y), the average number of hours of exercise per week (x_1), the cholesterol level (x_2), and the number of points that the individual's blood pressure exceeded the recommended value (x_3). A random sample of 40 individuals was selected. Suppose that a linear multiple regression model is appropriate for the analysis. The following table gives the values of the constant and the multiple regression coefficients, as well as their respective standard deviations (standard errors): Below are some of the findings:

ANOVA

	df	*SS*	*MS*	*F*
Regression		939.00		
		3,227.00		
		4,166.00		

	Coefficients	*Standard Error*	*t-Stat*
Constant	55.80	11.80	
X_1	1.790	0.440	
X_2	-0.021	0.011	
X_3	-0.016	0.014	

a. Interpret the meaning of the regression coefficient of X_2 in the context of the problem.

b. Is there evidence at the 5% significance level that the age at death is related to any of these explanatory (independent) variables?

c. Test the model's overall significance. Use $\alpha = 1\%$

d. Is there enough evidence at the 1% significance level to infer that the average number of hours of exercise per week and the age at death are linearly related?

e. Predict the age at death of a person who on average spends 8 hours on exercise per week, has a cholesterol level of 0.5, and whose blood pressure exceeds the recommended value by 10 points.

8. A random sample of 12 first year Concordia University alumni revealed that they donated the following amounts to their alma mater in 2009: \$210 \$200 \$170 \$180 \$200 \$190 \$180 \$200 \$220 \$160 \$180 \$200

   Assume that the donations in any year by the first year alumni are normally distributed.

   a. Estimate with 95% confidence the mean amount donated in 2009.

   b. What sample size would be needed to estimate the mean amount donated to within \$10 at 99% confidence.

c. Estimate with 95% confidence the proportion of alumni who donated more than \$200.

d. What sample size would be needed to estimate the proportion in part c) to within 1% with 95% confidence?

9. A controversial issue in sports is the use of the "instant replay" for making decisions on plays that are extremely close or hard to call by an official. A survey of players in each of four professional sports was conducted, asking them if they felt "instant replays" should be used to decide close or controversial calls. The results are as follows:

Use of Instant Replay

	Favour	Oppose
Football	22	2
Baseball	18	6
Basketball	15	26
Soccer	3	10

In testing to see whether opinion with respect to the use of instant replays is independent of sport, a table of expected frequencies is to be calculated.

a. Prepare the necessary Contingency Table.

b. In this table, the expected number of professional baseball players opposing the use of instant replays is equal to what value?

c. Test whether type of sports played (Football, Baseball, Basketball and Soccer) is independent of players opinion (in favour or oppose) at 5% significance level.

d. In this case what type of error do you make?

10. Many computer buyers have discovered that they can save a considerable amount by purchasing a PC from a mail-order company. A recent article claims that one can save $850 on average. To test this claim, 15 customers were contacted and asked how much they had saved by purchasing through the mail. The mean and standard deviation were, respectively, $835 and $30.
    a. Conduct an appropriate test for the claim at α=0.01 test.

    b. What are the Type I and Type II errors in the context of this problem? Which type of error you are risking?

    c. Calculate the p-value.

    d. What are the requirements for the test?