Elementary Probability and Statistics
Final Exam
June 6, 2011

↑ Student name and ID number ↑

Instructor: Bjørn Kjos-Hanssen

Disclaimer: It is essential to write legibly and *show your work*. If your work is absent or illegible, and at the same time your answer is not perfectly correct, then no partial credit can be awarded. Completely correct answers which are given without justification may receive little or no credit.

During this exam, you are not permitted to use notes, or books, nor to collaborate with others. You are allowed to use a calculator.

Problem	Score
1	/5
2	/6
3	/10
4	/10
5	/14
6	/9
7	/9
8	/12
Total	/75
times 100/75=	/100

Score: _______

Problem 1. [5 points] Suppose a store sells boxes of cream puffs. Each box contains 12 cream puffs, and each cream puff is either chocolate cream or vanilla cream. The number of chocolate cream puffs in a box has a mean of 4 and a standard deviation of 2.6. If you buy 7 boxes of cream puffs, what is the standard deviation of the number of vanilla cream puffs that you will get? You may assume that the numbers of vanilla cream puffs in different boxes are independent of one another.

Solution: Since the number of vanilla is a constant minus the number of chocolate, we are just looking for the SD of the number of chocolates for $n = 7$, where $\sigma = 2.6$ and $\mu = 4$. This is $\sigma\sqrt{n} = 2.6\sqrt{7} = 6.87$.

Score: _______

Problem 2. [6pts]

All the 143 Android phone visits to the web site *math.hawaii.edu* last month were done using either the default web browser or the Opera Mini browser.

- Four visits used Opera Mini. 1 of these 4 was a *new visit* (the other three being returning visits).
- 139 visits used the default browser. 30 of these were *new visits*.

a) [3 points] How many percent of all visits using Android phones were new visits?

b) [3 points] What percentage of new visits used the default browser?

Solution:

$$\frac{1+30}{143} = 21.7\%. \tag{a}$$

$$30/31 = 96.8\%. \tag{b}$$

Score: _______

Problem 3. Suppose that houses in La Jolla area are sold at a rate of 1.02 per day, and that on average, 13.3% of the houses sold are built in the first half of 1963 or earlier (we will call such houses "old").

Real estate agent Sally has noticed that the numbers of houses, old and new, and the numbers of buyers and sellers in the market, are very large compared to the number of sales that typically occur in a month. Therefore she adopts the following mathematical modeling assumptions: Ages of houses sold are independent of one another, and the number of sales, and the time until the next sale, are independent across time periods. Based on these assumptions answer (a)-(d) below.

a) [3 points] Find the probability that exactly 1 of the next 7 houses sold will be "old".

b) [3 points] Find the probability that exactly 8 houses will be sold in the next (7-day) week.

c) [2 points] What is the probability that it will be at least a 7-day week (from now) before the next house is sold?

d) [2 points] Suppose no houses are sold in April. What is the probability that no house will be sold in the first 7-day week of May?

Solution: (a) Binomial random variable X with $n = 7$ and $p = .133$. Then

$$\mathbb{P}(X = 1) = \binom{7}{1} p(1-p)^6 = 7(.133)(.867)^6 = .395.$$

(b) Let now X be Poisson with $\lambda = 7 \cdot (1.02) = 7.14$ being the rate at which houses are sold per week. Then

$$\mathbb{P}(X = 8) = e^{-\lambda}\lambda^8/8! = 0.1328.$$

(c) Let now X be an exponential random variable with $\lambda = 7.14$. Then

$$\mathbb{P}(X \geq 1) = 1 - \mathbb{P}(X < 1) = 1 - \int_0^1 \lambda e^{-\lambda x} dx = 1 - (1 - e^{-\lambda}) = e^{-(7.14)} = 0.00079 = .08\%$$

(d) Same as (c), .08%, by the memoryless property of the exponential distribution.

Score: _______

Problem 4. Suppose you read in the newspaper that 65% of men with mustaches (facial hair above the lips) also have a beard (facial hair below the lips). To test your theory, you somehow draw a simple random sample of 10 men having a mustache. As it turns out, you observe that none (zero) of these 10 men also has a beard.

a) [6 points] What can you conclude about whether or not the newspaper article you read was accurate? Make sure to state your hypotheses clearly, show how you calculated your test statistic, give the p-value (or an interval containing the p-value, if that is the best you can do with your tables), and write a clear conclusion. You may use a significance level of .05.

Let $p_0 = 65\%$, let p be the overall mean proportion among men with mustaches of those that also have a beard. We let the null hypothesis H_0 state that $p = p_0$ and the alternative hypothesis H_A state that $p \neq p_0$. (It is tempting but slightly wrong to let H_A be $p < p_0$, because initially the number $p_0 = 65\%$ was not stated as being high or low.) We can do a one-proportion z-test.

$$Z = \frac{p - p_0}{\sqrt{p_0(1-p_0)/n}} = \frac{0 - .65}{\sqrt{(.65)(.35)/10}}$$

$$= \frac{-65}{\sqrt{(65)(35)/10}} = \frac{-13}{\sqrt{(13)(7)/10}} = -\sqrt{130/7} = -4.3$$

According to the Table Z, the p-value is 0.0000^+, i.e., less than $0.0001 = 10^{-4}$. [Note that if we used a binomial rather than normal random variable, the probability would be $(.35)^{10} = .28 \cdot 10^{-4}$ so we are probably on the right track.]

b) [2 points] Explain what a Type I error would mean in the context of this problem.

A Type I error would mean rejecting the null hypothesis even though it is true, in other words, concluding that the newspaper article was wrong even though actually it was right.

c) [2 points] Explain what the power of the test means in the context of this problem.

The power of the test is the probability that we will reject the null hypothesis, in the case where the null hypothesis is false. That is, the probability in the case where the newspaper article is wrong, that we will be able to determine that using our test. [This probability will depend on what p actually is; if it is very close to p_0 the power is smaller for a fixed n.]

Score: _______

Problem 5. [1 point per question.]

Suppose we have some data $(x_1, y_1), (x_2, y_2), (x_3, y_3)$ given by $(0,0), (1,1), (2,3)$.

In the questions below, show how you are using a formula and end up giving your answer in numerical form; part (a) is shown as an example. Consult the Useful Formulas sheet as needed.

a) Find $\overline{x}$. Solution: $\overline{x} = \frac{x_1+x_2+x_3}{3} = \frac{0+1+2}{3} = 1$.

b) Find $\overline{y}$. Solution: $\overline{y} = \frac{0+1+3}{3} = \frac{4}{3} = 1.33$.

c) Find s_x.

$$= \sqrt{\frac{(0-1)^2 + (1-1)^2 + (2-1)^2}{2}} = 1.$$

d) Find s_y.

$$= \sqrt{\frac{(0-\frac{4}{3})^2 + (1-\frac{4}{3})^2 + (3-\frac{4}{3})^2}{2}} = \frac{1}{3}\sqrt{\frac{16+1+25}{2}} = \frac{1}{3}\sqrt{21} = \sqrt{7/3} = 1.53.$$

e) Find the correlation coefficient r.

$$= \frac{1}{2} \cdot \frac{1}{1 \cdot \sqrt{7/3}} \left((0-1)(0-\frac{4}{3}) + (1-1)(1-\frac{4}{3}) + (2-1)(3-\frac{4}{3}) \right)$$

$$= \frac{\sqrt{3}}{2 \cdot \sqrt{7}} \left(\frac{4}{3} + 0 + \frac{5}{3}) \right) = \frac{3\sqrt{3}}{2\sqrt{7}} = \sqrt{\frac{27}{28}} = 98.1\%$$

(f) Find the slope estimate b_1.

$$= \frac{r s_y}{s_x} = \sqrt{\frac{27 \cdot 7}{28 \cdot 3}} = \sqrt{\frac{9}{4}} = \frac{3}{2}$$

(g) Find the y-intercept estimate b_0.

$$= \overline{y} - b_1 \overline{x} = \frac{4}{3} - \frac{3}{2} = -\frac{1}{6}$$

Score: _______

(h) Find $s = s_e$.

At this point we have the regression line

$$\hat{y} = \frac{3}{2}x - \frac{1}{6}.$$

This gives

$$s_e = \sqrt{\frac{1}{n-2}\left((0-(-\frac{1}{6}))^2 + (1-\frac{4}{3})^2 + (3-(3-\frac{1}{6}))^2\right)}$$

$$= \sqrt{\frac{1}{6^2} + \frac{1}{3^2} + \frac{1}{6^2}} = \frac{1}{3}\sqrt{\frac{1}{4} + 1 + \frac{1}{4}} = \frac{\sqrt{3}}{3\sqrt{2}} = \frac{1}{\sqrt{6}} = .408$$

Note the residuals are $+1/3$ and $-1/6$ and $-1/6$, and this is slightly bigger in absolute value.

(i) Find the standard error of the slope estimate, $\mathrm{SE}(b_1)$. This is

$$\frac{s_e}{s_x\sqrt{n-1}} = \frac{1}{\sqrt{6}\sqrt{2}} = \frac{1}{2\sqrt{3}} = .29$$

(j) Find the value of the random variable T.

$$T = b_1/SE(b_1) = 2\sqrt{3}\cdot 3/2 = 3\sqrt{3} = 5.2.$$

(k) Find a 95% confidence interval for the slope β_1.

The number of df is $n-2 = 3-2 = 1$. The margin of error is

$$t^*_{n-2}SE(b_1) = 12.706/(2\sqrt{3}) = 3.67$$

so the confidence interval is

$$b_1 \pm \mathrm{ME} = [1.5 - 3.67, 1.5 + 3.67] = [-2.17, +5.17].$$

(l) Find the standard error for the mean response for the x-value $x^* = 1/2$ (denoted by $\mathrm{SE}(\hat{\mu}_y)$ in the Useful Formulas sheet).

$$SE(\hat{\mu}_y) = \sqrt{\frac{s_e^2}{n} + (x^* - \overline{x})^2 SE(b_1)^2} = \sqrt{\frac{1}{18} + \frac{1}{4}\cdot\frac{1}{12}} = \frac{1}{\sqrt{6}}\sqrt{\frac{1}{3} + \frac{1}{8}}$$

$$= \frac{1}{\sqrt{6}}\sqrt{\frac{8+3}{24}} = \frac{\sqrt{11}}{12} = .276$$

Score: ________

(m) Find the standard error for the predicted response for $x^* = 1/2$ (denoted SE($\hat{y}$) in the Useful Formulas sheet).

$$SE(\hat{y}) = \sqrt{s_e^2 + \frac{s_e^2}{n} + (x^* - \overline{x})^2 SE(b_1)^2} = \sqrt{\frac{3+1}{18} + \frac{1}{4} \cdot \frac{1}{12}} = \frac{1}{\sqrt{6}}\sqrt{\frac{4}{3} + \frac{1}{8}}$$

$$= \frac{1}{\sqrt{6}}\sqrt{\frac{32+3}{24}} = \frac{\sqrt{35}}{12} = .493$$

(n) Find a 95% confidence interval for the mean response for $x^* = 1/2$. The predicted value is $\frac{3}{2} \cdot \frac{1}{2} - \frac{1}{6} = \frac{7}{12} = .583$. Now the margin of error is (since the df is still $n - 2 = 1$)

$$12.706 \cdot \mathrm{SE}(\hat{\mu}_y) = 12.706(.276) = 3.5$$

so the confidence interval is $[.583 - 3.5, .583 + 3.5] \approx [.6 - 3.5, .6 + 3.5] = [-2.9, 4.1]$.

(o) Find a 95% prediction interval for the response for $x^* = 1/2$. The predicted value is still $\frac{7}{12}$ and now the margin of error is

$$12.706 \cdot \mathrm{SE}(\hat{y}) = 12.706(.493) = 6.3$$

so the prediction interval is $[.6 - 6.3, .6 + 6.3] = [-5.7, 6.9]$.

Score: _______

Additional space for answers to Problem 5.

Score: _______

Problem 6. Suppose 121 gamblers in Las Vegas are chosen at random, and their lifetime winnings or losses have an average of -\$4,700 (a loss of \$4,700) and a standard deviation of \$43,000.

a) [6 pts] Find a 99% confidence interval for the average winning or loss of all gamblers in Las Vegas.

b) [3 pts] Do you think approximately 99 percent of gamblers in Las Vegas have lifetime winnings in the interval that you found in part a)? Explain.

Solutions: (a) The confidence interval is centered at -4,700 and has a margin of error of $t^* s/\sqrt{n} = 2.617 \cdot 43,000/\sqrt{121} = 10,230.1$.

(b) No, the confidence interval is not a "prediction interval"; it is not true that 99 percent are that close to the average, only that we are 99 percent certain that the average is in the interval.

Score: _______

Problem 7. (9 pts) While several operating systems and web browsers are in use, here we will restrict attention to two operating systems (Windows and Mac) and two browsers (Firefox and Chrome); so we will assume that everybody is using either Windows or Mac, and either Firefox or Chrome.

The number of visits to the web site *math.hawaii.edu* using one of these operating systems and one of these browsers in May 2011 was as follows.

Number of visits	Mac	Windows
Firefox	538	1,788
Chrome	290	1,126

Conduct a χ^2 test of the hypothesis that the choice of browser is independent of the choice of operating system.

Solution: We widen the table:

Number of visits	Mac	Windows	TOTAL
Firefox	538	1,788	2,326
Chrome	290	1,126	1,416
TOTAL	828	2,914	3,742

We have the following expected cell counts:

Number of visits	Mac	Windows	TOTAL
Firefox	828 * 2326 / 3742 = 514.67878	2914*2326 / 3742 = 1811.3	2,326
Chrome	828* 1416 / 3742 = 313.3	2914 * 1416 / 3742 = 1102.67878	1,416
TOTAL	828	2,914	3,742

Differences between observations and expected counts with expected:

$$538 - 514.7 = 1788 - 1811.3 = 290 - 313.3 = 1126 - 1102.7 = 23.3$$

Squared: $23.3212186^2 = 543.879$. Sum of squared differences divided by expected counts:

$$543.879\left(\frac{1}{514.7} + \frac{1}{1811.3} + \frac{1}{313.3} + \frac{1}{1102.7}\right) = 3.586$$

X^2 for $(2-1)(2-1) = 1$ degree of freedom. Since 3.586 is between 2.706 and 3.841, the p-value is between 5% and 10%. Thus, we cannot reject the hypothesis of independence.

Score: _______

Problem 8. We wish to determine whether professors (currently working) have shorter last names, on average, than their doctoral advisers (who we assume are retired, so there is no overlap between professors and advisers). We have the following data, in the format (professor's last name length, adviser's last name length):

$(6,6),(6,9),(7,8),(9,6),(6,8),(6,6),(5,5),(6,4),(6,8),(5,5),(6,7),(4,7),(5,7),(6,8),(7,7),(6,8).$

(a) [5 points] Draw a histogram to check whether the differences (professor's last name minus that professor's adviser's last name) are approximately normally distributed.

The differences are 0, -3, -1, 3, -2, 0, 0, 2, -2, 0, -1, -3, -2, -2, 0, -2. Counts:

-3	xx
-2	xxxxx
-1	xx
0	xxxxx
1	
2	x
3	x

Looking at this sideways, it looks consistent with a normal distribution, considering that the sample size is only 16.

Score: _______

Problem 8, continued. The following facts can be calculated from the data (but you are not asked to do so): The professors' last names have a standard deviation of 1.1, the advisers' last names have a standard deviation of 1.4, and the differences have a standard deviation of 1.7. The average professor last name length is 6.0, the average adviser last name length is 6.8, and the average difference is -0.8.

(b) [7 points] Is there strong evidence that professors have shorter last names than their advisers?

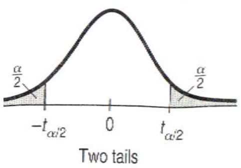
Justify your answer by conducting an appropriate hypothesis test at significance level .05. Make sure to give the p-value for your test, or an interval containing the p-value if that is the best you can do with your tables.

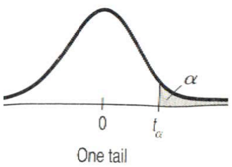
To use the paired t-test we should check that the differences are approximately normally distributed, which they are by part (a). The negative of the differences have mean $\bar{d} = +0.8$ and standard deviation $s_d = 1.7$. Therefore the test statistic should be

$$T = \bar{d}/(s_d/\sqrt{n}) = \sqrt{16} \cdot 0.8/1.7 = 4 \cdot 8/17 = 32/17 = 1.88.$$

The degrees of freedom is $n - 1 = 15$ and we are using a one-sided test since the question asked about "shorter" rather than "different". We see from Table T that a T value of 1.753 would give p-value 5% and a T value of 2.131 would give a p-value of 2.5%. Thus we are just barely able to reject the null hypothesis that professors and their advisers have the same length last names on average. But one might argue that a one-sided test is unnatural, and a two-sided test should have been done, in which case we could not reject H_0. We could also not reject at significance level 1%, for example. Further observation is probably required.

Table T
Values of t_α

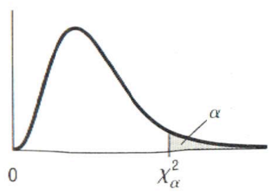




Two-tail probability	0.20	0.10	0.05	0.02	0.01	
One-tail probability	0.10	0.05	0.025	0.01	0.005	
df						df
1	3.078	6.314	12.706	31.821	63.657	1
2	1.886	2.920	4.303	6.965	9.925	2
3	1.638	2.353	3.182	4.541	5.841	3
4	1.533	2.132	2.776	3.747	4.604	4
5	1.476	2.015	2.571	3.365	4.032	5
6	1.440	1.943	2.447	3.143	3.707	6
7	1.415	1.895	2.365	2.998	3.499	7
8	1.397	1.860	2.306	2.896	3.355	8
9	1.383	1.833	2.262	2.821	3.250	9
10	1.372	1.812	2.228	2.764	3.169	10
11	1.363	1.796	2.201	2.718	3.106	11
12	1.356	1.782	2.179	2.681	3.055	12
13	1.350	1.771	2.160	2.650	3.012	13
14	1.345	1.761	2.145	2.624	2.977	14
15	1.341	1.753	2.131	2.602	2.947	15
16	1.337	1.746	2.120	2.583	2.921	16
17	1.333	1.740	2.110	2.567	2.898	17
18	1.330	1.734	2.101	2.552	2.878	18
19	1.328	1.729	2.093	2.539	2.861	19
20	1.325	1.725	2.086	2.528	2.845	20
21	1.323	1.721	2.080	2.518	2.831	21
22	1.321	1.717	2.074	2.508	2.819	22
23	1.319	1.714	2.069	2.500	2.807	23
24	1.318	1.711	2.064	2.492	2.797	24
25	1.316	1.708	2.060	2.485	2.787	25
26	1.315	1.706	2.056	2.479	2.779	26
27	1.314	1.703	2.052	2.473	2.771	27
28	1.313	1.701	2.048	2.467	2.763	28
29	1.311	1.699	2.045	2.462	2.756	29
30	1.310	1.697	2.042	2.457	2.750	30
32	1.309	1.694	2.037	2.449	2.738	32
35	1.306	1.690	2.030	2.438	2.725	35
40	1.303	1.684	2.021	2.423	2.704	40
45	1.301	1.679	2.014	2.412	2.690	45
50	1.299	1.676	2.009	2.403	2.678	50
60	1.296	1.671	2.000	2.390	2.660	60
75	1.293	1.665	1.992	2.377	2.643	75
100	1.290	1.660	1.984	2.364	2.626	100
120	1.289	1.658	1.980	2.358	2.617	120
140	1.288	1.656	1.977	2.353	2.611	140
180	1.286	1.653	1.973	2.347	2.603	180
250	1.285	1.651	1.969	2.341	2.596	250
400	1.284	1.649	1.966	2.336	2.588	400
1000	1.282	1.646	1.962	2.330	2.581	1000
∞	1.282	1.645	1.960	2.326	2.576	∞
Confidence levels	80%	90%	95%	98%	99%	

Figure 1: Areas under t distribution curves.

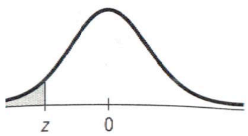
Table X
Values of χ^2_α



Right-tail probability / df	0.10	0.05	0.025	0.01	0.005
1	2.706	3.841	5.024	6.635	7.879
2	4.605	5.991	7.378	9.210	10.597
3	6.251	7.815	9.348	11.345	12.838
4	7.779	9.488	11.143	13.277	14.860
5	9.236	11.070	12.833	15.086	16.750
6	10.645	12.592	14.449	16.812	18.548
7	12.017	14.067	16.013	18.475	20.278
8	13.362	15.507	17.535	20.090	21.955
9	14.684	16.919	19.023	21.666	23.589
10	15.987	18.307	20.483	23.209	25.188
11	17.275	19.675	21.920	24.725	26.757
12	18.549	21.026	23.337	26.217	28.300
13	19.812	22.362	24.736	27.688	29.819
14	21.064	23.685	26.119	29.141	31.319
15	22.307	24.996	27.488	30.578	32.801
16	23.542	26.296	28.845	32.000	34.267
17	24.769	27.587	30.191	33.409	35.718
18	25.989	28.869	31.526	34.805	37.156
19	27.204	30.143	32.852	36.191	38.582
20	28.412	31.410	34.170	37.566	39.997
21	29.615	32.671	35.479	38.932	41.401
22	30.813	33.924	36.781	40.290	42.796
23	32.007	35.172	38.076	41.638	44.181
24	33.196	36.415	39.364	42.980	45.559
25	34.382	37.653	40.647	44.314	46.928
26	35.563	38.885	41.923	45.642	48.290
27	36.741	40.113	43.195	46.963	49.645
28	37.916	41.337	44.461	48.278	50.994
29	39.087	42.557	45.722	59.588	52.336
30	40.256	43.773	46.979	50.892	53.672
40	51.805	55.759	59.342	63.691	66.767
50	63.167	67.505	71.420	76.154	79.490
60	74.397	79.082	83.298	88.381	91.955
70	85.527	90.531	95.023	100.424	104.213
80	96.578	101.879	106.628	112.328	116.320
90	107.565	113.145	118.135	124.115	128.296
100	118.499	124.343	129.563	135.811	140.177

Figure 2: Areas under χ^2 distribution curves.

Areas under the standard Normal curve

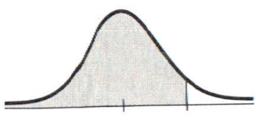


0.09	0.08	0.07	0.06	0.05	0.04	0.03	0.02	0.01	0.00	z
									0.0000†	−3.9
0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	−3.8
0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	−3.7
0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0002	0.0002	−3.6
0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	−3.5
0.0002	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	−3.4
0.0003	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0005	0.0005	0.0005	−3.3
0.0005	0.0005	0.0005	0.0006	0.0006	0.0006	0.0006	0.0006	0.0007	0.0007	−3.2
0.0007	0.0007	0.0008	0.0008	0.0008	0.0008	0.0009	0.0009	0.0009	0.0010	−3.1
0.0010	0.0010	0.0011	0.0011	0.0011	0.0012	0.0012	0.0013	0.0013	0.0013	−3.0
0.0014	0.0014	0.0015	0.0015	0.0016	0.0016	0.0017	0.0018	0.0018	0.0019	−2.9
0.0019	0.0020	0.0021	0.0021	0.0022	0.0023	0.0023	0.0024	0.0025	0.0026	−2.8
0.0026	0.0027	0.0028	0.0029	0.0030	0.0031	0.0032	0.0033	0.0034	0.0035	−2.7
0.0036	0.0037	0.0038	0.0039	0.0040	0.0041	0.0043	0.0044	0.0045	0.0047	−2.6
0.0048	0.0049	0.0051	0.0052	0.0054	0.0055	0.0057	0.0059	0.0060	0.0062	−2.5
0.0064	0.0066	0.0068	0.0069	0.0071	0.0073	0.0075	0.0078	0.0080	0.0082	−2.4
0.0084	0.0087	0.0089	0.0091	0.0094	0.0096	0.0099	0.0102	0.0104	0.0107	−2.3
0.0110	0.0113	0.0116	0.0119	0.0122	0.0125	0.0129	0.0132	0.0136	0.0139	−2.2
0.0143	0.0146	0.0150	0.0154	0.0158	0.0162	0.0166	0.0170	0.0174	0.0179	−2.1
0.0183	0.0188	0.0192	0.0197	0.0202	0.0207	0.0212	0.0217	0.0222	0.0228	−2.0
0.0233	0.0239	0.0244	0.0250	0.0256	0.0262	0.0268	0.0274	0.0281	0.0287	−1.9
0.0294	0.0301	0.0307	0.0314	0.0322	0.0329	0.0336	0.0344	0.0351	0.0359	−1.8
0.0367	0.0375	0.0384	0.0392	0.0401	0.0409	0.0418	0.0427	0.0436	0.0446	−1.7
0.0455	0.0465	0.0475	0.0485	0.0495	0.0505	0.0516	0.0526	0.0537	0.0548	−1.6
0.0559	0.0571	0.0582	0.0594	0.0606	0.0618	0.0630	0.0643	0.0655	0.0668	−1.5
0.0681	0.0694	0.0708	0.0721	0.0735	0.0749	0.0764	0.0778	0.0793	0.0808	−1.4
0.0823	0.0838	0.0853	0.0869	0.0885	0.0901	0.0918	0.0934	0.0951	0.0968	−1.3
0.0985	0.1003	0.1020	0.1038	0.1056	0.1075	0.1093	0.1112	0.1131	0.1151	−1.2
0.1170	0.1190	0.1210	0.1230	0.1251	0.1271	0.1292	0.1314	0.1335	0.1357	−1.1
0.1379	0.1401	0.1423	0.1446	0.1469	0.1492	0.1515	0.1539	0.1562	0.1587	−1.0
0.1611	0.1635	0.1660	0.1685	0.1711	0.1736	0.1762	0.1788	0.1814	0.1841	−0.9
0.1867	0.1894	0.1922	0.1949	0.1977	0.2005	0.2033	0.2061	0.2090	0.2119	−0.8
0.2148	0.2177	0.2206	0.2236	0.2266	0.2296	0.2327	0.2358	0.2389	0.2420	−0.7
0.2451	0.2483	0.2514	0.2546	0.2578	0.2611	0.2643	0.2676	0.2709	0.2743	−0.6
0.2776	0.2810	0.2843	0.2877	0.2912	0.2946	0.2981	0.3015	0.3050	0.3085	−0.5
0.3121	0.3156	0.3192	0.3228	0.3264	0.3300	0.3336	0.3372	0.3409	0.3446	−0.4
0.3483	0.3520	0.3557	0.3594	0.3632	0.3669	0.3707	0.3745	0.3783	0.3821	−0.3
0.3859	0.3897	0.3936	0.3974	0.4013	0.4052	0.4090	0.4129	0.4168	0.4207	−0.2
0.4247	0.4286	0.4325	0.4364	0.4404	0.4443	0.4483	0.4522	0.4562	0.4602	−0.1
0.4641	0.4681	0.4721	0.4761	0.4801	0.4840	0.4880	0.4920	0.4960	0.5000	−0.0

†For $z \le -3.90$, the areas are 0.0000 to four decimal places.

Figure 3: Areas under the standard normal curve.

Areas under the standard Normal curve



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.9	1.0000†									

†For $z \geq 3.90$, the areas are 1.0000 to four decimal places.

Figure 4: Areas under the standard normal curve, continued.

Useful Formulas

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \qquad s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2} \qquad r = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{s_x}\right)\left(\frac{y_i - \bar{y}}{s_y}\right)$$

$$\hat{y} = b_0 + b_1 x \qquad b_1 = \frac{r s_y}{s_x} \qquad b_0 = \bar{y} - b_1\bar{x}$$

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B) \qquad P(A \text{ and } B) = P(A)P(B|A)$$

$$P(B|A) = \frac{P(A \text{ and } B)}{P(A)} \qquad P(A) = P(B)P(A|B) + P(B^c)P(A|B^c)$$

$$P(B|A) = \frac{P(B)P(A|B)}{P(B)P(A|B) + P(B^c)P(A|B^c)}$$

	Discrete case	**Continuous case**
$P(a \le X \le b)$	$\sum_{i:a\le x_i \le b} P(X = x_i)$	$\int_a^b f(x)\,dx$
$\mu = E[X]$	$\sum_i x_i P(X = x_i)$	$\int_{-\infty}^{\infty} x f(x)\,dx$
$\text{Var}(X)$	$\sum_i (x_i - \mu)^2 P(X = x_i)$	$\left(\int_{-\infty}^{\infty} x^2 f(x)\,dx\right) - \mu^2$
$\text{SD}(X)$	$\sqrt{\text{Var}(X)}$	$\sqrt{\text{Var}(X)}$

Distribution	**probability function or density**	$E[X]$	$\text{SD}(X)$
Binomial(n, p)	$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$ for $k = 0, 1, ..., n$	np	$\sqrt{np(1-p)}$
Poisson(λ)	$P(X = k) = e^{-\lambda}\lambda^k/k!$ for $k = 0, 1, \ldots$	λ	$\sqrt{\lambda}$
Geometric(p)	$P(X = k) = (1-p)^{k-1}p$ for $k = 1, 2, ...$	$1/p$	$\sqrt{1-p}/p$
Uniform(a, b)	$f(x) = 1/(b-a)$ for $a < x < b$	$(a+b)/2$	$(b-a)/\sqrt{12}$
Exponential(λ)	$f(x) = \lambda e^{-\lambda x}$ for $x > 0$	$1/\lambda$	$1/\lambda$
Normal(μ, σ^2)	$f(x) = (2\pi\sigma^2)^{-1/2} e^{-(x-\mu)^2/2\sigma^2}$	μ	σ

$$E[X + c] = E[X] + c \qquad E[cX] = cE[X] \qquad E[X + Y] = E[X] + E[Y].$$

$$\text{Var}(X + c) = \text{Var}(X) \qquad \text{Var}(cX) = c^2\text{Var}(X)$$

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) \text{ if } X \text{ and } Y \text{ are independent.}$$

$$E[\bar{X}] = \mu \qquad \text{SD}(\bar{X}) = \sigma/\sqrt{n} \qquad E[\hat{p}] = p \qquad \text{SD}(\hat{p}) = \sqrt{\frac{p(1-p)}{n}}$$

Test	ME for Interval	Test Statistic	df
One-proportion z-test	$z^*\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$	$Z = \dfrac{p - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$	N/A
One-sample t-test	$t^*\dfrac{s}{\sqrt{n}}$	$T = \dfrac{\bar{X} - \mu_0}{s/\sqrt{n}}$	$n-1$
Paired t-test	$t^*\dfrac{s_d}{\sqrt{n}}$	$T = \dfrac{\bar{d}}{s_d/\sqrt{n}}$	$n-1$
Two-sample t-test	$t^*\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}$	$T = \dfrac{\bar{X} - \bar{Y}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$	$\min\{n_1 - 1, n_2 - 1\}$
Regression slope	$t^* SE(b_1)$	$T = \dfrac{b_1}{SE(b_1)}$	$n-2$
Chi-square test	N/A	$\chi^2 = \sum \dfrac{(\text{Obs - Exp})^2}{\text{Exp}}$	see below

Chi-square test for goodness-of-fit: df = number of categories - 1

Chi-square test for independence/homogeneity: df = (rows - 1)(columns - 1)

Expected counts in chi-square test for independence/homogeneity: $\dfrac{\text{Row Total} \times \text{Column Total}}{\text{Table Total}}$

$$SE(\hat{\mu}_y) = \sqrt{\frac{s^2}{n} + (x^* - \bar{x})^2 SE(b_1)^2} \qquad SE(\hat{y}) = \sqrt{s^2 + \frac{s^2}{n} + (x^* - \bar{x})^2 SE(b_1)^2}$$

$$SE(b_1) = \frac{s}{s_x\sqrt{n-1}}$$