Examination Number:____________

Last Name_____________First__________

Sign the Honor Pledge Below_____________ PID #________________________________

__________________________________ Write Your Section Number here:_________

University of North Carolina
Economics 400: Economic Statistics & Econometrics
Second Midterm Examination

Prof. B. Turchi April 4, 2019

General Instructions: Answer all 3 questions (with multiple parts) on this examination, writing your answers on the exam paper itself. Use the scratch page at the end for any extra work, if necessary. Sign the Honor Pledge above. Express all answers to a precision of at least 3 decimal points. Show your work to be eligible for partial credit. **Be sure to note that tables and formulas are on the last 4 pages of the exam.**

1. On February 11, 2019, the *News & Observer* published an article about the high failure rates of new teachers in North Carolina on the newly adopted **Pearson Mathematics Competency** test. Only 65 percent of the first-time test takers passed the exam as compared to a much higher pass rate on the previous math test. Responding to complaints, the NC Board of Education has included an alternative test, the **Praxis** test as an option for test-takers. Assuming that test takers in 2019 are randomly assigned to either the Pearson or the Praxis test, and that there are approximately 1,200 people taking each exam, sample (without replacement) 50 test takers from each group.

(a) (5 points) Theoretically, what is the strictly most appropriate distribution to use when computing probabilities from each of these two samples? Why?

The Hypergeometric distribution because we’re sampling without replacement.

(b) (5 points) In performing hypothesis tests on each of these two samples what probability distributions <u>could</u> you use? Which one <u>would</u> you use? Why?

I could use either the hypergeometric, the binomial, or the normal distribution. The hypergeometric distribution is too complicated to use. I could use the binomial because, even though we’re sampling without replacement, the samples are small enough, relative to the population size, (n is less than 5 percent of N) to allow its use. However, I <u>would</u> use the normal approximation to the binomial because both rules of thumb [np>5 and n(1-p)>5] are met.

(c) (10 points)The State would like at least 70 percent of test-takers to pass the Pearson exam. In the sample of 50 test takers, 31 passed the exam. Can you reject at the 5% significance level the null hypothesis that 70 percent or <u>more</u> of test takers passed the exam? Show your work and explain your answer.

This is a hypothesis test of a population proportion. We have a sample of 50 test-takers 31 of whom passed the test. We want to test the null hypothesis that the pass rate is 70 percent or more vs. the alternative that less than 70 percent passed.

$$H_0 : P \geq 70\%$$
$$H_a : P < 70\%$$
$$\hat{P} = \frac{31}{50} = 0.62$$
$$P_0 = 0.70$$

$$\text{Test statistic: } z = \frac{\hat{P} - P_0}{\sqrt{\frac{P_0(1-P_0)}{n}}} = \frac{0.62 - 0.70}{\sqrt{\frac{0.70(1-0.70)}{50}}} = \frac{-0.08}{0.06481} = -1.23443$$

Critical Value (Left Tail Test): $\alpha = 0.05: \quad -1.645$

$\therefore$ Cannot reject the null hypothesis

(d) (10 points) The State would like at least 70 percent of test-takers to pass the Praxis exam. In the sample of 50 test takers, 39 passed the exam. Can you reject at the 5% significance level the null hypothesis that 70 percent or more of test takers passed the exam? Show your work and explain your answer.

This is a hypothesis test of a population proportion. We have a sample of 50 test-takers 39 of whom passed the test. We want to test the null hypothesis that the pass rate is 70 percent or more vs. the alternative that less than 70 percent passed.

$$H_0 : P \geq 70\%$$
$$H_a : P < 70\%$$
$$\hat{P} = \frac{39}{50} = 0.78$$
$$P_0 = 0.70$$

$$\text{Test statistic: } z = \frac{\hat{P} - P_0}{\sqrt{\frac{P_0(1-P_0)}{n}}} = \frac{0.78 - 0.70}{\sqrt{\frac{0.70(1-0.70)}{50}}} = \frac{0.08}{0.06481} = 1.23443$$

Critical Value (Left Tail Test): $\alpha = 0.05: \quad -1.645$

$\therefore$ Cannot reject the null hypothesis

(e) (10 points) From the two samples, it appears that the Praxis exam has a higher pass rate than the Pearson exam. Perform a hypothesis test in which you test the null hypothesis that, in the population, the two tests have the same pass rate versus the alternative hypothesis that the Pearson test is more difficult. Use the 5% significance level. Show your work and explain your answer.

This is a test comparing two population proportions from independent samples which are large enough so that we can assume the normal distribution to test the null hypothesis. We need to compute an estimate of the true population proportions which we will call the "pooled sample

$H_0 : P_{\Pr axis} = P_{Pearson}$

$H_a : P_{Pearson} < P_{\Pr axis}$

Pooled Sample Proportion: $\hat{P}_p = \frac{x_1 + x_2}{n_1 + n_2} = \frac{31+39}{50+50} = 0.70$

Test Statistic: $z = \frac{\hat{P}_1 - \hat{P}_2}{\sqrt{\hat{P}_p\left(1-\hat{P}_p\right)} \bullet \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{-0.16}{0.45826 \times 0.20} = \frac{-0.16}{0.09165} = -1.7457$

Critical Value (Left Tail Test): $\alpha = 0.05 : -1.645$

$\therefore$ Can reject the null hypothesis at the 5% level (p-value actually = 0.0404).

proportion."

2. (33 points) It is well known that life expectancies at birth (the number of years a newly born person can be expected to survive) vary significantly across countries. These variations seem to be related to variations in Gross Domestic Product per capita. Using data from the United Nations Human Development Indicators data set, I regressed life expectancy at birth ("life") versus GDP percapita ("gdp") for 179 countries, and the results are:

```
. regress life gdp

      Source |       SS           df       MS      Number of obs   =       179
-------------+----------------------------------   F(1, 177)       =    xxxxxx
       Model |  6654.46757         1  6654.46757   Prob > F        =    0.0000
    Residual |  11157.8378       177  63.0386318   R-squared       =    xxxxxx
-------------+----------------------------------   Adj R-squared   =    0.3700
       Total |  17812.3054       178  100.069131   Root MSE        =    7.9397

------------------------------------------------------------------------------
        life |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         gdp |   .0004385   .0000427    xxxxx   xxxxx      .0003543    .0005228
       _cons |   63.22705   .7872336    80.32   0.000      61.67348    64.78062
------------------------------------------------------------------------------
```

(a) The F-statistic for this regression is: _______, the null hypothesis it tests is:________________________________, and we can/cannot (circle one) reject the null hypothesis.

(b) The t-statistic for the slope coefficient is: __________, and the null hypothesis it tests is:____

(c) The p-value for the slope coefficient is (to 3 significant digits):______________, and we can/cannot (circle one) reject the null hypothesis.

(d) What is the R^2 for the regression?____

(e) What is the correlation between "life" and "gdp"?__________________________

(f) What is the standardized slope coefficient for this regression?________________

(g) Draw a causal diagram in the box below showing the relationship between "life" and "gdp" (include the error term). Do the variations in "gdp" cause significant variations in "life"? Why? Could the causation run in the opposite direction?

____
____
____

(viii) Name the three assumption about the error term that need to be true for the regression above to represent an unbiased view of the relationship between life expectancy and GDP per capita:

(1)

____
____

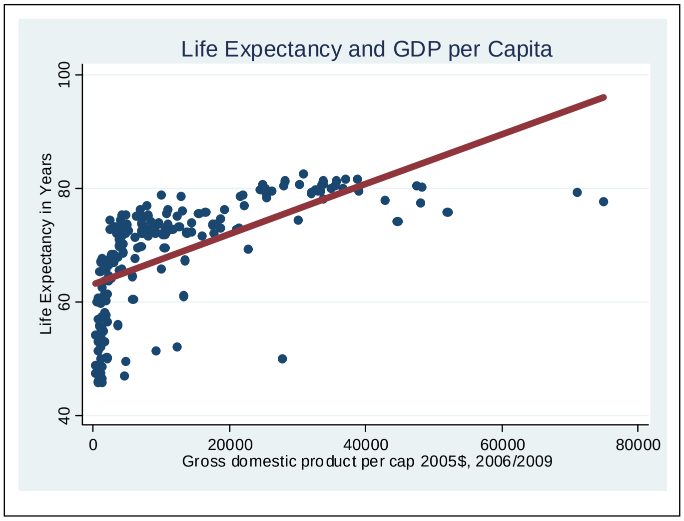
(2)

____
____

(3)

____
____

Below, find a scatter plot of life expectancy and GDP per capita superimposed on the regression line estimated above:



(ix) Does fitting a straight line through these data adequately represent the relationship between GDP per capita and life expectancy? Why?/Why not?

___

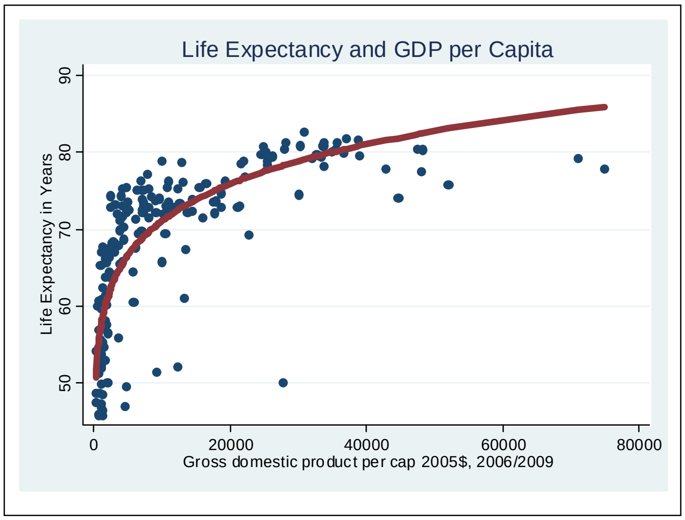
___

___

(x) What change to the regression above might have been used to produce the regresssion line below? Do you think this regression would have a higher/lower/identical (circle 1) R^2 as compared to the first regression, Why?

___

___



(xi) What *other* problem might make the original regression an inadequate representation of the true relationship between the variables? (name at least one problem)

____

____

____

Answers to Question 2:

```
. regress life gdp

      Source |       SS           df       MS      Number of obs   =       179
-------------+----------------------------------   F(1, 177)       =    105.56
       Model |  6654.46757         1  6654.46757   Prob > F        =    0.0000
    Residual |  11157.8378       177  63.0386318   R-squared       =    0.3736
-------------+----------------------------------   Adj R-squared   =    0.3700
       Total |  17812.3054       178  100.069131   Root MSE        =    7.9397

------------------------------------------------------------------------------
        life |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         gdp |   .0004385   .0000427    10.27   0.000     .0003543    .0005228
       _cons |   63.22705   .7872336    80.32   0.000     61.67348    64.78062
------------------------------------------------------------------------------
```

(i) F-statistic for this regression is: 105.56, the null hypothesis it tests is: H_0: F = 1, and we can/~~cannot~~ (circle one) reject the null hypothesis.

We get this by dividing the model mean squared error by the residual mean squared error: $\frac{6654.46757}{63.0386318} = 105.56$

(ii) The t-statistic for the slope coefficient is: 10.27, and the null hypothesis it tests is: $H_0 : \beta_{gdp} = 0$ That is, the slope coefficient = 0

The t-statistic is calculated as $\frac{\hat{\beta}}{s_{\hat{\beta}}} = \frac{0.0004385}{0.0000427} = 10.27$. It is also equal to the square root of the F-statistic.

(iii) The p-value for the slope coefficient is (to 3 significant digits): 0.000, and we can/~~cannot~~ (circle one) reject the null hypothesis.

The sutdent could look at the t-table and see that, at 3 significant digits the p-value is smaller than 0.000. Looking at the Prob>F statistic gives you the same conclusion.

(iv) What is the R^2 for the regression? 0.3736.

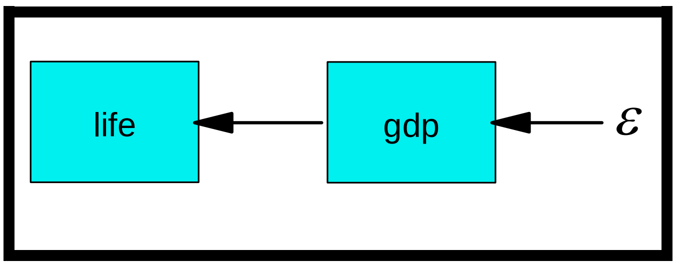
The R^2 is calculated by dividing the model sum of squares by the total sum of squares: $\frac{6654.46757}{17812.3054} = 0.3736$

(v) What is the correlation between "life" and "gdp"?0.6112 which is the square root of R^2.

(vi) What is the standardized slope coefficient for this regression: 0.6112

(vii) Draw a causal diagram in the box below showing the relationship between "life" and "gdp" (include the error term). Do the variations in "gdp" cause significant variations in "life"? Why? Could the causation run in the opposite direction?

Yes, based on the regression results (t-test on the Beta coefficient, or, equivalently, the hypothesis test on the entire regression (F-test). On the other hand, it could be the case that differences in life expectancy at birth cause differences in GDP per capita. High mortality could lower the productivity of the population.



(viii) Name the three assumption about the error term that need to be true for the regression above to represent an unbiased view of the relationship between life expectancy and GDP per capita:

(1) The error term needs to be independent of the independent variable, gdp.

(2) The expected value of the error term need to equal zero for all observations.

(3) The covariance of all error terms = 0 and the variance of all the error terms is the same

(ix) Does fitting a straight line through these data adequately represent the relationship between GDP per capita and life expectancy? Why?/Why not?

No, fitting a straight line through the scatter shows that it does not adequately represent the relationship between the two variables. The observations should be randomly distributed around the regression line. They are not. They do not appear to be independent of the independent variable. The slope coefficient is likely biased. Two issues arise, either together or individually:

1. The functional form of the relationship is not linear, but curvilinear.

2. There may be other, omitted, variables in the relationship that, if included, would lead to independence between the independent variable(s) and the error term. However, in that case we would have a multiple regression

(x) (a) What change to the regression above might have been used to produce the regresssion line below? (b) Do you think this regression would have a higher/~~lower/identical~~ (circle 1) R^2 as compared to the first regression, Why?

(a) We fit a curve through the scatter of data points. This curve leads to a much better fit with the scatter of data and it appears that the residuals have a much higher likelihood of being independent of the independent variable. The students don't have to say this, but I fit a double-log regression through the data (I could also have used a semi-log regression) and got the following results:

$$life = a \bullet gdp^{\beta} \Rightarrow \ln life = \ln a + \beta \bullet \ln gdp = 3.396 + 0.0943 \bullet \ln gdp$$

(b) Comparing the two diagrams, it is clear that the second regression would have a higher R^2 than the first. In fact, for the second regression we have:

$$R^2 = 0.60$$

Which is more than 60% greater than the original regression. Of course the students don't have the actual number, but comparison of the two plots should lead to the correct conclusion.

(xi) What *other* problem might make the original regression an inadequate representation of the true relationship between the variables? (name at least one problem)

(a) There could be omitted variables that are affecting the coefficient on *gdp*. This could lead to omitted variable bias.

3. Ever since shrimp fishing resumed in the the Gulf of Mexico after the oil spill catastrophe in 2010, concerns have been raised about the toxicity of the harvested shrimp. Being the disgusting bottom feeders that they are, are shrimp too toxic to eat? Using a measure of toxicity that has a mean permissible level of 50 units, health inspectors are testing random samples of shrimp for toxicity level. The data below are from the first sample taken from fishing area **A** off the coast of Louisiana. The question is, is the mean toxicity level low enough to allow the shrimp to be sold on the market?

(a) (9 points) Test the hypothesis that mean toxicity levels are at, or below the required level versus the alternative hypothesis that they aren't. Use significance level 0.05. Explain your answer, in particular, (i) what assumption(s) must you make about the data; (ii) what distribution should you use? Why? (iii) Can you reject the null hypothesis? Why/Why not? Explain your answer.

(i) (2 points) The sample is small enough so that we should assume that the sample has been drawn from a normal population.

```
     areaA
     +-------+         +-------+
  1. |   140 |    13. |     59 |
  2. |   102 |    14. |     27 |
  3. |    73 |    15. |    -15 |
  4. |    40 |        |-------|
  5. |    85 |    16. |     60 |
     |-------|    17. |     36 |
  6. |   153 |    18. |    104 |
  7. |   131 |    19. |     48 |
  8. |    57 |    20. |     34 |
  9. |    58 |        |-------|
 10. |    12 |    21. |     59 |
     |-------|    22. |     92 |
 11. |    40 |    23. |     62 |
 12. |    27 |    24. |     51 |
     +-------+         +-------+
```

(ii) (2 points) Because we have to estimate the population standard deviation from the sample data, we should use the t-distribution to test our null hypothesis.

(iii) (5 points) This is a standard t-test on a sample mean. The null and alternative hypotheses are:

$$H_0 : \mu \le 50$$
$$H_a : \mu > 50$$

The sample mean and standard deviation are:

$$\bar{x} = 63.95833 \text{ where } \bar{x} = \frac{\sum_{i=1}^{24} x_i}{24}$$

$$s = 40.45124 \text{ where } s = \sqrt{\frac{\sum_{i=1}^{24} (x_i - \bar{x})^2}{24-1}}$$

The t-statistic is:

$$t_{df=23} = \frac{\bar{x} - \mu_{H_0}}{\frac{s}{\sqrt{24}}} = 1.6905$$

The critical value for an alpha level = 0.05 and 23 degrees of freedom is : 1.714. So, the t-statistic lies inside the do not reject region and we cannot reject the null hypothesis.

(b) (5 points) Another catch, randomly sampled from Area B, off the coast of Mississippi, has the following values. **Sample Size = 24, Sample Mean Toxicity = 87.70833, Sample Standard Deviation = 42.37563.** Test this batch to see if, at the 0.05 significance level, one can reject the null hypothesis that mean toxicity levels are at, or below the required level, versus the alternative that they aren't. Can you reject the null hypothesis? Why?/Why not?

Here, we just do another hypothesis test with the values above substituted into the t-statistic formula:

$$t_{df=23} = \frac{\bar{x} - \mu_{H_0}}{\frac{s}{\sqrt{24}}} = \frac{87.70833 - 50}{\frac{42.37563}{\sqrt{24}}} = 4.3594$$

Here the critical value of 1.714 for *df=23* is far exceeded by the t-statistic and we can conclusively reject the null hypothesis tha toxicity levels are at or below the mandated level.

(c) (6 points) Using the two sample standard deviations you have available, test the hypothesis that the population variances from the two fishing areas are equal to each other versus the alternative that they're not. (i) What is the appropriate distribution to use for this hypothesis test? (ii) Show the formula for the test statistic and the the value of the test statistic you computed. (iii) Based on that value, and using the appropriate statistical table, can you reject the null hypothesis of equal population variances? Why?/Why not?

(i) (2 points) We need to use the F-distribution

(ii) (2 points)

$$F_{(23,23)} = \frac{s_B^2}{s_A^2} = 1.097$$

(iii) (2 points) This F-statistic (ratio) is so close to one, that it would be extremely hard to reject the null hypothesis that the two population variances are equal. Looking at the F-table for ((20 or 24),23) degrees of freedom, we see that the critical value for rejection of the null hypothesis is > 2.0, so we cannot reject the null hypothesis of equal population variances.

(d) (7 points) Fishermen from Mississippi claim that the toxicity levels of their shrimp are no different from those caught in Louisiana waters. Are they correct? Based on your answer to part (c) above, perform the appropriate hypothesis test of the difference between two means. (i) Using a two tail test, can you reject the null hypothesis at the <u>0.05 significance level?</u> Why?/Why not? (ii) Using a one-tail test, can you reject the null hypothesis at <u>the 0.01 significance level?</u> Why?/Why not?

(i) (4 points) We have two independent samples and we are not able to reject the null hypothesis that the two population standard deviations are equal. Hence, we need to run a two sample t-test assuming equal population variances. The t-test is as follows:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_p\sqrt{(1/n_1) + (1/n_2)}} = -1.9861$$

The pooled standard deviation is:

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} = 41.4246$$

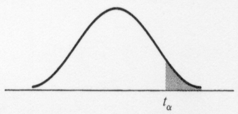
With *df=46 the critical values for a two tail test are plus/minus 1.96 so you just miss rejecting at the alpha=.05 level (p-value = 0.0530*

(ii) (3 points) *For the one tail hypothesis at alpha = .01 the critical value is -2.326, so, again you cannot reject at the alpha = 0.01 level. The p-value is 0.0265.*

The student may have incorrectly used the unpooled tests. In this case, the results are virtually Identical:

```
Two-sample t test with unequal variances
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
   areaA |      24    63.95833    8.257074    40.45124    46.87727    81.03939
   areaB |      24    87.70833    8.649889    42.37563    69.81467     105.602
---------+--------------------------------------------------------------------
combined |      48    75.83333    6.163575     42.7025    63.43382    88.23285
---------+--------------------------------------------------------------------
    diff |              -23.75    11.95825               -47.82212    .3221189
------------------------------------------------------------------------------
    diff = mean(areaA) - mean(areaB)                              t =  -1.9861
Ho: diff = 0                     Satterthwaite's degrees of freedom =   45.901

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0265         Pr(|T| > |t|) = 0.0530          Pr(T > t) = 0.9735
```



d.f.	$t_{.100}$	$t_{.050}$	$t_{.025}$	$t_{.010}$	$t_{.005}$	d.f.
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
inf.	1.282	1.645	1.960	2.326	2.576	inf.

Source: From "Table of Percentage Points of the *t*-Distribution." *Biometrika*, Vol. 32 (1941), p. 300. Reproduced by permission of the Biometrika Trustees.

$$\Pr\left(-t_{\alpha/2} < \frac{\bar{x}-\mu}{s/\sqrt{n}} < t_{\alpha/2}\right) = 1-\alpha$$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_p\sqrt{(1/n_1)+(1/n_2)}}$$

$$df = n_1 + n_2 - 2.$$

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

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

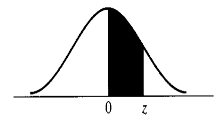
$$\Delta = \frac{\left[(s_1^2/n_1)+(s_2^2/n_2)\right]^2}{\frac{(s_1^2/n_1)^2}{n_1-1}+\frac{(s_2^2/n_2)^2}{n_2-1}}$$ rounded down.

$$\hat{p}_p = \frac{x_1+x_2}{n_1+n_2}$$

$$z = \frac{\hat{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

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

Normal Curve Areas



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

Source: This table is abridged from Table 1 of *Statistical Tables and Formulas*, by A. Hald (New York: John Wiley & Sons, Inc., 1952). Reproduced by permission of A. Hald and the publishers, John Wiley & Sons, Inc.

$$n = \frac{z_{\alpha/2}^2\sigma^2}{D^2}. \qquad n = \frac{z_{\alpha/2}^2}{4D^2}$$

Binomial Coefficients

n	$\binom{n}{0}$	$\binom{n}{1}$	$\binom{n}{2}$	$\binom{n}{3}$	$\binom{n}{4}$	$\binom{n}{5}$	$\binom{n}{6}$	$\binom{n}{7}$	$\binom{n}{8}$	$\binom{n}{9}$	$\binom{n}{10}$
0	1										
1	1	1									
2	1	2	1								
3	1	3	3	1							
4	1	4	6	4	1						
5	1	5	10	10	5	1					
0	1	6	15	20	15	6	1				
7	1	7	21	35	35	21	7	1			
8	1	8	28	56	70	56	28	8	1		
9	1	9	36	84	126	126	84	36	9	1	
10	1	10	45	120	210	252	210	120	45	10	1
11	1	11	55	165	330	462	462	330	165	55	11
12	1	12	66	220	405	792	924	792	495	220	66
13	1	13	78	286	715	1287	1716	1716	1287	715	286
14	1	14	91	364	1001	2002	3003	3432	3003	2002	1001
15	1	15	105	455	1365	3003	5005	6435	6435	5005	3003
16	1	16	120	560	1820	4368	8008	11440	12870	11440	8008
17	1	17	136	680	2380	6188	12376	19448	24310	24310	19448
18	1	18	153	816	3060	8568	18564	31824	43758	48620	43758
19	1	19	171	969	3876	11628	27132	50388	75582	92378	92378
20	1	20	190	1140	4845	15304	38760	77520	125970	167960	184756

If necessary, use the identity $\binom{n}{k} = \binom{n}{n-k}$.

$$f(\tilde{x}) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \quad -\infty < x < \infty$$

$$\sigma^2 = \frac{\sum_{i=1}^{N}(x_i-\mu)^2}{N}$$

$$p(x) = \frac{C_x^r C_{n-x}^{N-r}}{C_n^N} = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$$

Mean: $\mu = n\left(\frac{r}{N}\right)$

Variance: $\sigma^2 = n\left(\frac{r}{N}\right)\left(\frac{N-r}{N}\right)\left(\frac{N-n}{N-1}\right)$

Standard deviation: $\sigma = \sqrt{\sigma^2}$

$$f(x) = \begin{cases} \dfrac{1}{(b-a)}, & a \le x \le b \\ 0, & \text{otherwise} \end{cases}$$

$$\mu = \frac{1}{2}(b+a) \text{ and } \sigma = \frac{(b-a)}{\sqrt{12}}$$

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & \lambda > 0, x \ge 0 \\ 0, & \text{otherwise} \end{cases}$$

$$\mu = \frac{1}{\lambda} \text{ and } \sigma = \frac{1}{\lambda}$$

$$P(x \ge a) = e^{-\lambda a},\ a \ge 0 \text{ and } \lambda > 0$$

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{\text{all } k} P(B|A_k)P(A_k)}$$

$$F_{(n_A-1;n_B-1)} = \frac{s_A^2/\sigma_A^2}{s_B^2/\sigma_B^2}$$

$$\chi^2_{(n-1)} = \frac{n-1}{\sigma_0^2}s^2$$

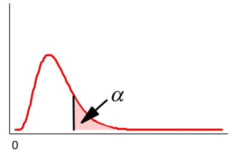
$$P(x) = \begin{cases} \dfrac{e^{-\lambda t}(\lambda t)^x}{x!}, & \text{for } x = 0,1,2,\ldots,\infty, \quad \lambda > 0, \\ 0, & \text{otherwise.} \end{cases}$$

λ = the mean number of events in a given segment of time $(t = 1)$

t = the length of a particular subsegment $(t \le 1)$

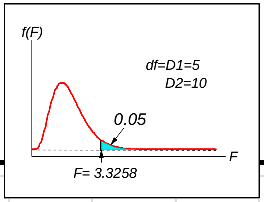
$E[x] = \mu_x = \lambda t$ = the expected number of events in one subsegment length t

Critical values of χ^2_α



df	$\chi^2_{0.995}$	$\chi^2_{0.99}$	$\chi^2_{0.975}$	$\chi^2_{0.95}$	$\chi^2_{0.90}$	$\chi^2_{0.10}$	$\chi^2_{0.05}$	$\chi^2_{0.025}$	$\chi^2_{0.001}$	$\chi^2_{0.005}$	df
1	0.000	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879	1
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597	2
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838	3
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860	4
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750	5
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548	6
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278	7
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955	8
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589	9
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188	10
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757	11
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300	12
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819	13
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319	14
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801	15
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267	16
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718	17
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156	18
19	6.844	7.633	8.907	10.117	11.651	27.204	30.143	32.852	36.191	38.582	19
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997	20
21	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401	21
22	8.643	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.290	42.796	22
23	9.260	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181	23
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.559	24
25	10.520	11.524	13.120	14.611	16.473	34.382	37.653	40.647	44.314	46.928	25
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290	26
27	11.808	12.879	14.573	16.151	18.114	36.741	40.113	43.195	46.963	49.645	27
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.994	28
29	13.121	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.336	29
30	13.787	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672	30
40	20.707	22.164	24.433	26.509	29.051	51.805	55.759	59.342	63.691	66.767	40
50	27.991	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490	50
60	35.534	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.381	91.955	60
70	43.275	45.442	48.758	51.739	55.329	85.527	90.531	95.023	100.424	104.213	70
80	51.172	53.540	57.153	60.391	64.278	96.578	101.879	106.628	112.328	116.320	80
90	59.196	61.754	65.647	69.126	73.291	107.565	113.145	118.135	124.115	128.296	90
100	67.328	70.065	74.222	77.930	82.358	118.499	124.343	129.563	135.811	140.177	100

F-Distribution Table: Upper 5% Probability (or 5% Area) under F-distribution Curve



			F-Table for alpha = 0.05																
/	df1=1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	inf
df2=1	161.4476	199.5	215.7073	224.5832	230.1619	233.986	236.7684	238.8827	240.5433	241.8817	243.906	245.9499	248.0131	249.0518	250.0951	251.1432	252.1957	253.2529	254.3144
2	18.5128	19	19.1643	19.2468	19.2964	19.3295	19.3532	19.371	19.3848	19.3959	19.4125	19.4291	19.4458	19.4541	19.4624	19.4707	19.4791	19.4874	19.4957
3	10.128	9.5521	9.2766	9.1172	9.0135	8.9406	8.8867	8.8452	8.8123	8.7855	8.7446	8.7029	8.6602	8.6385	8.6166	8.5944	8.572	8.5494	8.5264
4	7.7086	6.9443	6.5914	6.3882	6.2561	6.1631	6.0942	6.041	5.9988	5.9644	5.9117	5.8578	5.8025	5.7744	5.7459	5.717	5.6877	5.6581	5.6281
5	6.6079	5.7861	5.4095	5.1922	5.0503	4.9503	4.8759	4.8183	4.7725	4.7351	4.6777	4.6188	4.5581	4.5272	4.4957	4.4638	4.4314	4.3985	4.365
6	5.9874	5.1433	4.7571	4.5337	4.3874	4.2839	4.2067	4.1468	4.099	4.06	3.9999	3.9381	3.8742	3.8415	3.8082	3.7743	3.7398	3.7047	3.6689
7	5.5914	4.7374	4.3468	4.1203	3.9715	3.866	3.787	3.7257	3.6767	3.6365	3.5747	3.5107	3.4445	3.4105	3.3758	3.3404	3.3043	3.2674	3.2298
8	5.3177	4.459	4.0662	3.8379	3.6875	3.5806	3.5005	3.4381	3.3881	3.3472	3.2839	3.2184	3.1503	3.1152	3.0794	3.0428	3.0053	2.9669	2.9276
9	5.1174	4.2565	3.8625	3.6331	3.4817	3.3738	3.2927	3.2296	3.1789	3.1373	3.0729	3.0061	2.9365	2.9005	2.8637	2.8259	2.7872	2.7475	2.7067
10	4.9646	4.1028	3.7083	3.478	3.3258	3.2172	3.1355	3.0717	3.0204	2.9782	2.913	2.845	2.774	2.7372	2.6996	2.6609	2.6211	2.5801	2.5379
11	4.8443	3.9823	3.5874	3.3567	3.2039	3.0946	3.0123	2.948	2.8962	2.8536	2.7876	2.7186	2.6464	2.609	2.5705	2.5309	2.4901	2.448	2.4045
12	4.7472	3.8853	3.4903	3.2592	3.1059	2.9961	2.9134	2.8486	2.7964	2.7534	2.6866	2.6169	2.5436	2.5055	2.4663	2.4259	2.3842	2.341	2.2962
13	4.6672	3.8056	3.4105	3.1791	3.0254	2.9153	2.8321	2.7669	2.7144	2.671	2.6037	2.5331	2.4589	2.4202	2.3803	2.3392	2.2966	2.2524	2.2064
14	4.6001	3.7389	3.3439	3.1122	2.9582	2.8477	2.7642	2.6987	2.6458	2.6022	2.5342	2.463	2.3879	2.3487	2.3082	2.2664	2.2229	2.1778	2.1307
15	4.5431	3.6823	3.2874	3.0556	2.9013	2.7905	2.7066	2.6408	2.5876	2.5437	2.4753	2.4034	2.3275	2.2878	2.2468	2.2043	2.1601	2.1141	2.0658
16	4.494	3.6337	3.2389	3.0069	2.8524	2.7413	2.6572	2.5911	2.5377	2.4935	2.4247	2.3522	2.2756	2.2354	2.1938	2.1507	2.1058	2.0589	2.0096
17	4.4513	3.5915	3.1968	2.9647	2.81	2.6987	2.6143	2.548	2.4943	2.4499	2.3807	2.3077	2.2304	2.1898	2.1477	2.104	2.0584	2.0107	1.9604
18	4.4139	3.5546	3.1599	2.9277	2.7729	2.6613	2.5767	2.5102	2.4563	2.4117	2.3421	2.2686	2.1906	2.1497	2.1071	2.0629	2.0166	1.9681	1.9168
19	4.3807	3.5219	3.1274	2.8951	2.7401	2.6283	2.5435	2.4768	2.4227	2.3779	2.308	2.2341	2.1555	2.1141	2.0712	2.0264	1.9795	1.9302	1.878
20	4.3512	3.4928	3.0984	2.8661	2.7109	2.599	2.514	2.4471	2.3928	2.3479	2.2776	2.2033	2.1242	2.0825	2.0391	1.9938	1.9464	1.8963	1.8432
21	4.3248	3.4668	3.0725	2.8401	2.6848	2.5727	2.4876	2.4205	2.366	2.321	2.2504	2.1757	2.096	2.054	2.0102	1.9645	1.9165	1.8657	1.8117
22	4.3009	3.4434	3.0491	2.8167	2.6613	2.5491	2.4638	2.3965	2.3419	2.2967	2.2258	2.1508	2.0707	2.0283	1.9842	1.938	1.8894	1.838	1.7831
23	4.2793	3.4221	3.028	2.7955	2.64	2.5277	2.4422	2.3748	2.3201	2.2747	2.2036	2.1282	2.0476	2.005	1.9605	1.9139	1.8648	1.8128	1.757
24	4.2597	3.4028	3.0088	2.7763	2.6207	2.5082	2.4226	2.3551	2.3002	2.2547	2.1834	2.1077	2.0267	1.9838	1.939	1.892	1.8424	1.7896	1.733
25	4.2417	3.3852	2.9912	2.7587	2.603	2.4904	2.4047	2.3371	2.2821	2.2365	2.1649	2.0889	2.0075	1.9643	1.9192	1.8718	1.8217	1.7684	1.711
26	4.2252	3.369	2.9752	2.7426	2.5868	2.4741	2.3883	2.3205	2.2655	2.2197	2.1479	2.0716	1.9898	1.9464	1.901	1.8533	1.8027	1.7488	1.6906
27	4.21	3.3541	2.9604	2.7278	2.5719	2.4591	2.3732	2.3053	2.2501	2.2043	2.1323	2.0558	1.9736	1.9299	1.8842	1.8361	1.7851	1.7306	1.6717
28	4.196	3.3404	2.9467	2.7141	2.5581	2.4453	2.3593	2.2913	2.236	2.19	2.1179	2.0411	1.9586	1.9147	1.8687	1.8203	1.7689	1.7138	1.6541
29	4.183	3.3277	2.934	2.7014	2.5454	2.4324	2.3463	2.2783	2.2229	2.1768	2.1045	2.0275	1.9446	1.9005	1.8543	1.8055	1.7537	1.6981	1.6376
30	4.1709	3.3158	2.9223	2.6896	2.5336	2.4205	2.3343	2.2662	2.2107	2.1646	2.0921	2.0148	1.9317	1.8874	1.8409	1.7918	1.7396	1.6835	1.6223
40	4.0847	3.2317	2.8387	2.606	2.4495	2.3359	2.249	2.1802	2.124	2.0772	2.0035	1.9245	1.8389	1.7929	1.7444	1.6928	1.6373	1.5766	1.5089
60	4.0012	3.1504	2.7581	2.5252	2.3683	2.2541	2.1665	2.097	2.0401	1.9926	1.9174	1.8364	1.748	1.7001	1.6491	1.5943	1.5343	1.4673	1.3893
120	3.9201	3.0718	2.6802	2.4472	2.2899	2.175	2.0868	2.0164	1.9588	1.9105	1.8337	1.7505	1.6587	1.6084	1.5543	1.4952	1.429	1.3519	1.2539
inf	3.8415	2.9957	2.6049	2.3719	2.2141	2.0986	2.0096	1.9384	1.8799	1.8307	1.7522	1.6664	1.5705	1.5173	1.4591	1.394	1.318	1.2214	1