1 One-Sided Testing

In a two-sided test, we hypothesize that $H_0 : \mu = \mu_0$ and look for evidence that it's wrong; such evidence would be a t-statistic too big in either the positive or negative direction, expressed as $H_1 : \mu \neq \mu_0$.

When we do a one-sided test, we are only concerned with whether the true mean is either below or above our guess, but not both. For instance, suppose we think that μ is greater than μ_0 and we want to test this guess. The claim being tested becomes the *alternative* hypothesis. So we test, say at 5% significance,

$$H_0 : \mu \leq \mu_0,$$

$$H_1 : \mu > \mu_0.$$

We again assume that the null is true. We reject the null if we find strong enough evidence against the null, in favor of the alternative. Based on the specification, that evidence would be seen as a value of $\bar{x}$ that is "far enough" above μ_0, in other words, if $\bar{x} - \mu_0$ is very positive.

We quantify "far enough" by again using the t-statistic,

$$t \equiv \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \sim T(n-1).$$

But again, we only reject the null if t is too far *positive*, and hence we only look at the right-tail of the distribution. Hence we put all 5% of the rejection region into the right-tail. Thus our critical value is $t_{n-1,0.05}$ = `qt(1-0.05,n-1)` in R. We reject the null hypothesis if $t > t_{n-1,0.05}$. In other words, the rejection region is $(t_{n-1,0.05}, \infty)$.

If instead we think that μ is less than μ_0, the test becomes

$$H_0 : \mu \geq \mu_0,$$

$$H_1 : \mu < \mu_0.$$

In this setup, evidence against the null is when $\bar{x}$ is "far enough" below μ_0. Thus, if the t-statistic is too far *negative*, then we reject the null. This means we are only considering the left-tail of the distribution, in which we put all 5% of the test significance. The critical value is therefore $-t_{n-1,0.05}$ = `qt(0.05,n-1)` in R. We reject the null hypothesis if $t < -t_{n-1,0.05}$. In other words, the rejection region is $(-\infty, -t_{n-1,0.05})$.

Rule of Thumb: Put the hypothesis that contains the equality as the null hypothesis.

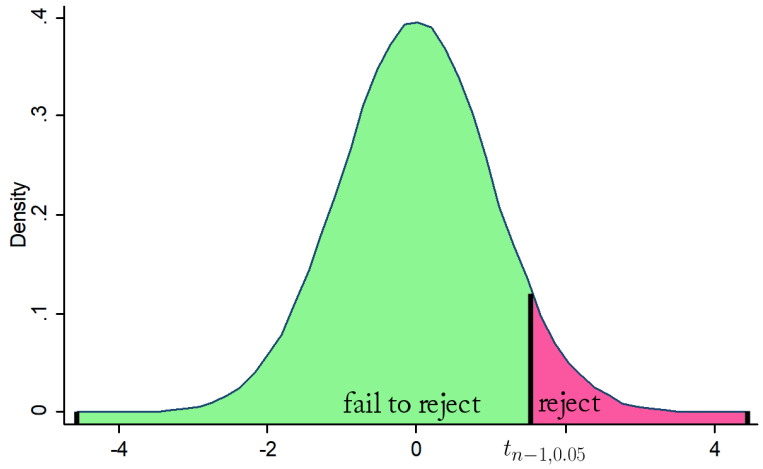

FIGURE 1: $t_{n-1,0.05}$ is the number such that 5% of the mass of the $T(n-1)$ distribution falls above it. If $H_0 : \mu \leq \mu_0$ is true, then it is unlikely that our test statistic will fall above $t_{n-1,0.05}$, in which case we reject the null.

2 Difference in Means Testing

First let me say that there are two difference of means tests. One assumes that the two groups have equal variances; the other does not. Here I do the version where variances are assumed unequal (which is typically the case in reality but not necessarily in a classroom). You can find the case where variances are assumed equal in slides on Canvas, but I omit it.

Suppose we are interested in two groups and how their means, μ_1 and μ_2, differ. We calculate sample means $\bar{x}_1$ and $\bar{x}_2$ as well as sample variances s_1^2 and s_2^2. We hypothesize that the difference in means is Δ_0; in practice we will often hypothesize that the difference is $\Delta_0 = 0$. Thus our null hypothesis is $H_0 : \mu_1 - \mu_2 = \Delta_0$. Hence we test

$$H_0 : \mu_1 - \mu_2 = \Delta_0,$$

$$H_1 : \mu_1 - \mu_2 \neq \Delta_0.$$

Suppose group 1 has sample size n_1 and group 2 has sample size n_2, not necessarily equal. The test statistic is

$$t \equiv \frac{(\bar{x}_1 - \bar{x}_2) - \Delta_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim T(n_1 + n_2 - 2).$$

The reason we subtract 2 for degrees of freedom is because we are testing with respect to two variables, μ_1 and μ_2. From here, the testing procedure proceeds in the usual way.

I illustrated the case of a two-sided test, but you should be able to extend this to a one-sided test as well.

3 One Proportion Testing

For individual i, let $x_i = 1$ for a "successful" event and $x_i = 0$ for a "failure" event. For example, earning a degree might be the successful event, dropping out would therefore be the failure event. The sample proportion of individuals who succeeded is the typical mean, now denoted $p \equiv (\sum_{i=1}^{n} x_i)/n$. Think of p as being an estimate of the true population proportion of successes, π.

Because there are only two possibilities for x_i, we have to use special techniques and formulas. In particular, the standard error of estimate p is given by

$$\text{se}(p) = \sqrt{\frac{p(1-p)}{n}}.$$

Furthermore, sample sizes in proportions analysis are typically large. Large enough, in fact, that the standard normal distribution is typically used instead of $T(n-1)$. Thus we do not use a t-statistic but instead the z-statistic given by

$$z \equiv \frac{p - \pi_0}{\sqrt{\pi_0(1-\pi_0)/n}} \sim \mathcal{N}(0,1)\,,$$

where π_0 is our hypothesized value for the true proportion of successful events.

A two-sided proportion test would be of the form

$$H_0 : \pi = \pi_0,$$

$$H_1 : \pi \neq \pi_0.$$

At 5% significance, we reject the null hypothesis when $|z| > z_{0.025}$, where in R you use $z_{0.025} =$ `qnorm(1-0.025)` or is otherwise found on a normal table.

Note that for this analysis to be valid, we require that $n\pi_0 \geq 10$ and $n(1-\pi_0) \geq 10$. And again, I illustrated the case of a two-sided test, but you should be able to extend this to a one-sided test as well.

4 Two Proportions Testing

Suppose we have two different population proportions, π_A and π_B. We want to see whether the proportions are the same or not. We sample n_A times for group A and find y_A successes; we sample n_B times for group B and find y_B successes. Hence we find estimates

$$p_A = \frac{y_A}{n_A}, \quad p_B = \frac{y_B}{n_B},$$

and the total proportion of successes is

$$\bar{p} = \frac{y_A + y_B}{n_A + n_B}.$$

Our test is of the form

$$H_0 : \pi_A - \pi_B = \Delta_0,$$

$$H_1 : \pi_A - \pi_B \neq \Delta_0.$$

In practice, we will often have $\Delta_0 = 0$, that is, we'll test if there is any difference. We use test statistic

$$z \equiv \frac{(p_A - p_B) - \Delta_0}{\sqrt{\bar{p}(1-\bar{p})\left(\frac{1}{n_A} + \frac{1}{n_B}\right)}} \sim \mathcal{N}(0,1)\,.$$

For this analysis to be valid, we require that $n\pi_0 \geq 5$ and $n(1-\pi_0) \geq 5$ for both n_A and n_B. I illustrated the case of a two-sided test, but you should be able to extend this to a one-sided test as well, deja vu, yes.

5 Chi-Square Distribution

A **chi-square** random variable, denoted χ^2, is a sum of squared standard normal random variables. Many test statistics have chi-square distribution, so we need to know about it. It has one parameter, the degrees of freedom k, and as such it is usually denoted $\chi^2(k)$. On quizzes and exams, we'll use a chi-square table (which is available on Canvas).

Suppose we have 20 degrees of freedom. We want to know critical value such that 5% of the area underneath the chi-square curve lies to the right of it. Then we go to the table, look at the row with 20 degrees of freedom and the column with 0.05, which gives 31.410. Express this number as $\chi^2_{20,0.05} =$ `qchisq(1-0.05,20)` in R.

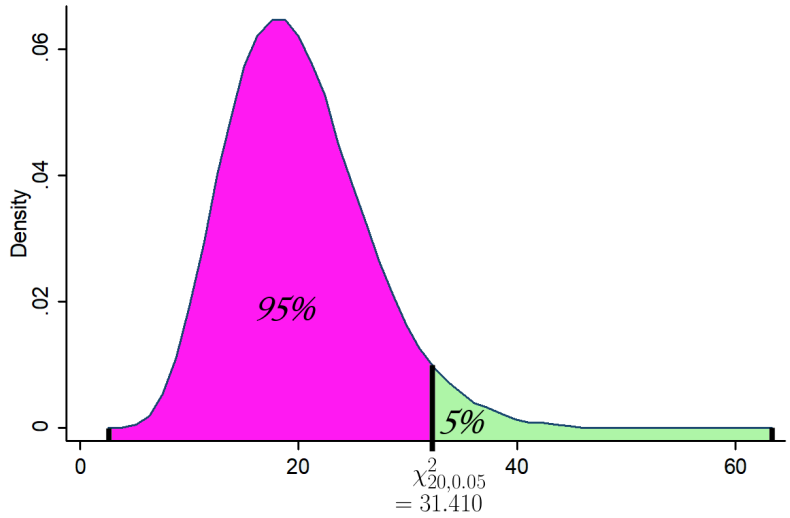

FIGURE 2: $\chi^2_{20,0.05}$ is the number such that 5% of the mass of $\chi^2(20)$ distribution falls above it.

6 Variance Testing

We usually do not know the true population variance σ^2, so we have to estimate it with

$$s^2 \equiv \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2.$$

But this is an estimation using a sample, and hence it has some uncertainty to it. We seek to quantify that uncertainty. For a two-sided test, we perform the test

$$H_0 : \sigma^2 = \sigma_0^2,$$
$$H_1 : \sigma^2 \neq \sigma_0^2,$$

where σ_0^2 is the hypothesized value for σ^2. The relevant test statistic is

$$\chi^2 \equiv \frac{(n-1)s^2}{\sigma_0^2} \sim \chi^2(n-1),$$

where the distribution is valid if either the population is normally distributed or if $n > 30$. We reject the null hypothesis if χ^2 is in the rejection region, which we now define.

Suppose we are testing at the 10% significance level. As usual, we chop the significance level in half for each tail. Problem is, the $\chi^2(n-1)$ distribution is not symmetric. Thus we must calculate two critical values to determine the rejection region. The values

can be found using the χ^2 table on Canvas. For instance, suppose that $n = 10$. Then we have $n - 1 = 9$ degrees of freedom. We want to find

$$\chi^2_{9,0.05} = \text{the number such that 0.05 of the area is to the right of it,}$$
$$\chi^2_{9,0.95} = \text{the number such that 0.95 of the area is to the right of it.}$$

These two numbers are visualized below.

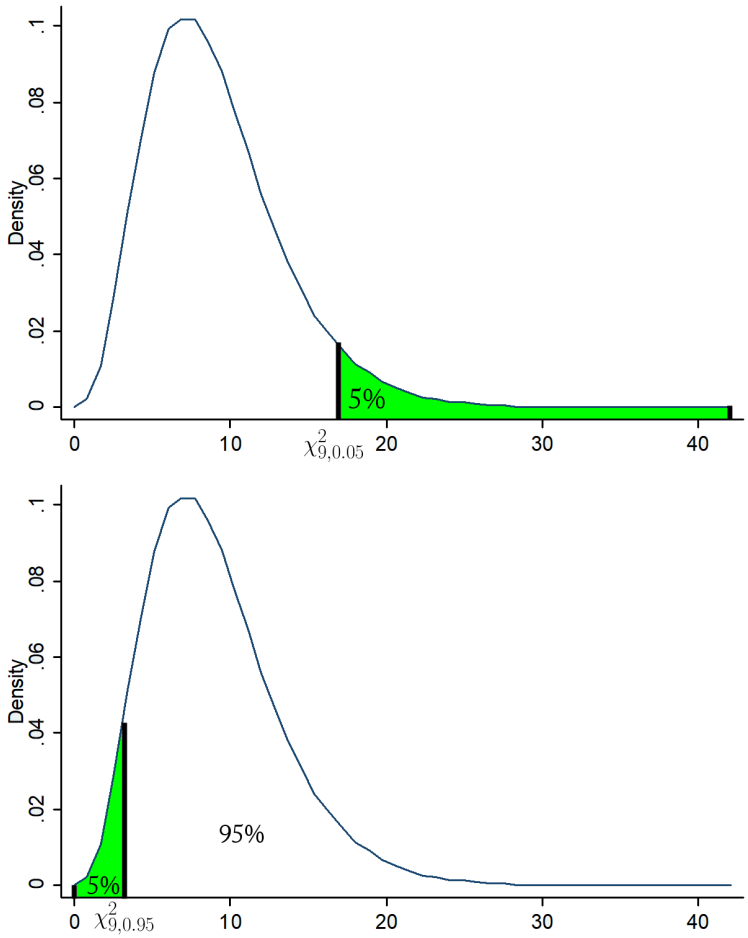

FIGURE 3: $\chi^2_{9,0.05}$ (top) is the number such that 5% of the mass of the $\chi^2(9)$ distribution falls above it, and $\chi^2_{9,0.95}$ (bottom) is the number such that 95% of the mass of the $\chi^2(9)$ distribution falls above it. Any χ^2 statistic in the green regions warrants rejecting the null.

So using the table, go to the 9th row, since that corresponds to 9 degrees of freedom. Then look at the columns for 0.95 and 0.05. Those are the critical values we seek.

Chi-Square (χ^2) Distribution

Degrees of Freedom	Area to the Right of Critical Value							
	0.99	0.975	0.95	0.90	0.10	0.05	0.025	0.01
1	—	0.001	0.004	0.016	2.706	3.841	5.024	6.635
2	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210
3	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345
4	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277
5	0.554	0.831	1.145	1.610	9.236	11.071	12.833	15.086
6	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812
7	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475
8	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090
9	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666
10	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209
11	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725
12	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217
13	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688
14	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141
15	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578

FIGURE 4: $\chi^2_{9,0.05} = 16.919$ is the number such that 5% of the mass of $\chi^2(9)$ distribution falls above it, and $\chi^2_{9,0.95} = 3.325$ is the number such that 95% of the mass of $\chi^2(9)$ distribution falls above it.

These values can be found in R using $\chi^2_{9,0.05} =$ `qchisq(0.05, 9, lower.tail = FALSE)` and $\chi^2_{9,0.95} =$ `qchisq(0.95, 9, lower.tail = FALSE)`.

7 F-Distribution

The ***F*-distribution** has two different arguments for two different degrees of freedom, so we denote it $F(v_1, v_2)$. What exactly v_1 and v_2 are will become clear once we start testing with it. Since we have to specify two degrees of freedom, it's difficult to condense a comprehensive F-distribution onto a single page. Hence we use an F table that contains only numbers for a 5% right-tail (therefore making it applicable to a one-sided test at 5% significance or a two-sided test with 10% significance, since in the latter we split the 10% into both tails). The table is currently available on Canvas.

Suppose $v_1 = 3$ and $v_2 = 15$, and we want to find the critical value of $F(3, 15)$ distribution such that 5% of the data falls to the right of it. Then we look at the row corresponding to $v_2 = 15$ degrees of freedom, and the column corresponding to $v_1 = 3$ degrees of freedom, which gives 3.287. Express this number as $F_{3,15,0.05} =$ `qf(1-0.05, 3, 15)` in R.

8 Difference in Variations Testing

Suppose we have two groups. Group A has sample size n and group B has sample size m. We calculate two different sample variances for each group, s_A^2 and s_B^2. We want to test

if one has true population variance greater than the other. Suppose that the samples give $s_A^2 > s_B^2$. Then let us test $H_0 : \sigma_A^2 \leq \sigma_B^2$ against $H_1 : \sigma_A^2 > \sigma_B^2$ at 5% significance.

Since our F-table only gives us right tails, we want to formulate the question in such a way that the test statistic leads to rejection if it is too far in the right tail. So reformulate the test as

$$H_0 : \frac{\sigma_A^2}{\sigma_B^2} \leq 1,$$

$$H_1 : \frac{\sigma_A^2}{\sigma_B^2} > 1,$$

where we use test statistic

$$F \equiv \frac{s_A^2}{s_B^2} \sim F(n-1, m-1).$$

Thus, if we have evidence in favor of the alternative, then F will be greater than 1. If F is sufficiently greater than 1, then we reject the null.

Rule of Thumb: Put the group with the larger sample variance in the numerator. This will ensure that we do a right-tailed test at 5% significance, which is all our F-table allows.

The critical value we use for the rejection rule is found by looking at the F-table. The numerator degrees of freedom is $v_1 = n - 1$, which is found on the columns; and the denominator degrees of freedom is $v_2 = m - 1$, which is found on the rows.

9 Errors in Conclusion

Since we are never 100% confident in our conclusions, it is possible that we reject a null hypothesis even when it is true; and also possible that we fail to reject a null hypothesis even when it is false. We employ the following terminology to discuss such scenarios.

- *Type I Error*: Rejection of a true null hypothesis (false positive)
- *Type II Error*: Failing to reject a false null hypothesis (false negative)

The *size* of a test is the probability of mistakenly rejecting a true null. The *power* of a test is the probability of correctly rejecting a false null. A test is said to have significance level α if its size is less than or equal to α. In many cases (and all of *our* cases), the size and significance level of a test are equal.

10 Examples

Example 1

Last Halloween, I ate 84 Starburst candies. However, not all econ grad students have an unquenchable need for Starburst. I don't know how many Starburst econ grad students ate on average, but I'm interested in finding out the variance in Starburst consumption last Halloween because I want to know just how out of hand my Starburst habit was.

I tracked down the Starburst consumption for $n = 31$ econ grad students. The average was $\bar{x} = 22$ and the variance was $s^2 = 14$. Someone told me that the true variance in Starburst consumption among econ grad students is actually $\sigma_0^2 = 8$. I think they're full of crap and I want to demonstrate how wrong they are with 95% confidence. Can I?

Solution. The test being performed is

$$H_0 : \sigma^2 = 8,$$
$$H_1 : \sigma^2 \neq 8.$$

The test statistic is

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{(30)14}{8} = 52.5.$$

The two critical values can be found on the χ^2 table, row 30, the columns with 0.975 and 0.025. The lower critical value is $\chi^2_{30,0.975} = 16.799$, the upper critical value $\chi^2_{30,0.025} = 46.979$. Since the test statistic is beyond the interval $[19.799, 46.997]$, which means it is in the rejection region, we reject the null hypothesis. Thus, I can tell that person how full of crap they are at 5% significance[1]: "If your guess was true, then there's a less than 5% chance that I'd have actually calculated $s^2 = 14$. So you're probably wrong."

[1] "Full of crap at 5% significance" is not standard statistical jargon.

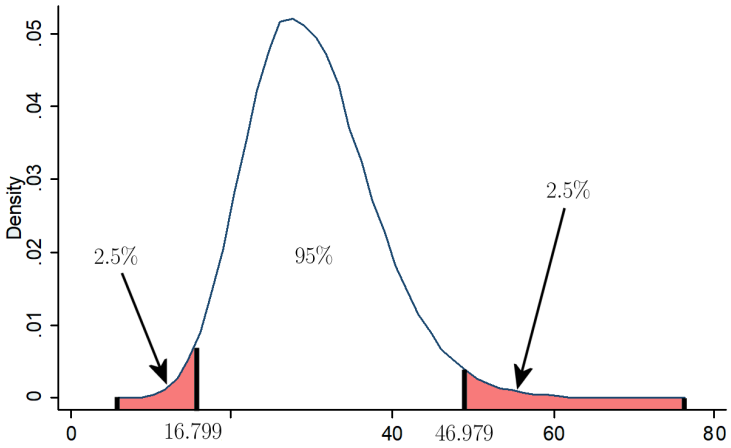

FIGURE 5: If the null is true, then there's a less than 5% chance of seeing a χ^2 statistic in the red regions. Since we found $\chi^2 = 52.5$, we reject the null.

Example 2

I also tracked down the Starburst consumption for $m = 21$ political science grad students. Their average was $\bar{x}_P = 28$ and the variance was $s_P^2 = 11$, compared to $\bar{x}_E = 22$ and $s_E^2 = 14$ for econ grad students. Someone told me that the true variance in Starburst consumption among political science grad students is lower than that among econ grad students. Test this claim at 5% significance.

Solution. This is a one-sided test (which will always be the case for our F-table questions), so the claim (with the strict inequality) becomes the alternative hypothesis. Rephrase "variance in Starburst consumption among political science grad students is lower" as "variance in Starburst consumption among econ grad students is higher." The test is

$$H_0 : \frac{\sigma_E^2}{\sigma_P^2} \leq 1,$$

$$H_1 : \frac{\sigma_E^2}{\sigma_P^2} > 1,$$

where we use test statistic

$$F \equiv \frac{s_E^2}{s_P^2} \sim F(n-1, m-1),$$

such that $n-1$ is the numerator (econ) degrees of freedom, and $m-1$ is the denominator (polisci) degrees of freedom. Thus we reject the null in favor of the alternative if we find a test statistic sufficiently larger than 1 (which is consistent with the claim that $\sigma_E^2 > \sigma_P^2$).

Rule of Thumb: Put the group with the larger sample variance in the numerator. This will ensure that we do a right-tailed test at 5% significance, which is all our F-table allows.

Our test statistic here is

$$F = \frac{14}{11} \approx 1.273.$$

Since we are testing this at 5% significance, we can use the F-table on Canvas. We look at numerator column $v_1 = n - 1 = 30$ and denominator column $v_2 = m - 1 = 20$ and find the critical value of 2.039. Our test statistic is below the critical value, hence we fail to reject the null: we have insufficient evidence to claim with 95% confidence that $\sigma_E^2 > \sigma_P^2$.

F-Distribution (p=0.95)

Use for one-tail tests at significance level 5% or two-tail tests at significance level 10%.

v_1	1	2	3	4	5	6	7	8	9	10	11	12	15	20	25	30	40	50	100	∞	v_1
v_2																					v_2
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.0	243.9	245.9	248.0	249.3	250.1	251.1	251.8	253.0	254.3	1
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.40	19.41	19.43	19.45	19.46	19.46	19.47	19.48	19.49	19.50	2
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.76	8.74	8.70	8.66	8.63	8.62	8.59	8.58	8.55	8.53	3
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.94	5.91	5.86	5.80	5.77	5.75	5.72	5.70	5.66	5.63	4
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.70	4.68	4.62	4.56	4.52	4.50	4.46	4.44	4.41	4.36	5
6	5.987	5.143	4.757	4.534	4.387	4.284	4.207	4.147	4.099	4.060	4.027	4.000	3.938	3.874	3.835	3.808	3.774	3.754	3.712	3.669	6
7	5.591	4.737	4.347	4.120	3.972	3.866	3.787	3.726	3.677	3.637	3.603	3.575	3.511	3.445	3.404	3.376	3.340	3.319	3.275	3.230	7
8	5.318	4.459	4.066	3.838	3.688	3.581	3.500	3.438	3.388	3.347	3.313	3.284	3.218	3.150	3.108	3.079	3.043	3.020	2.975	2.928	8
9	5.117	4.256	3.863	3.633	3.482	3.374	3.293	3.230	3.179	3.137	3.102	3.073	3.006	2.936	2.893	2.864	2.826	2.803	2.756	2.707	9
10	4.965	4.103	3.708	3.478	3.326	3.217	3.135	3.072	3.020	2.978	2.943	2.913	2.845	2.774	2.730	2.700	2.661	2.637	2.588	2.538	10
11	4.844	3.982	3.587	3.357	3.204	3.095	3.012	2.948	2.896	2.854	2.818	2.788	2.719	2.646	2.601	2.570	2.531	2.507	2.457	2.404	11
12	4.747	3.885	3.490	3.259	3.106	2.996	2.913	2.849	2.796	2.753	2.717	2.687	2.617	2.544	2.498	2.466	2.426	2.401	2.350	2.296	12
13	4.667	3.806	3.411	3.179	3.025	2.915	2.832	2.767	2.714	2.671	2.635	2.604	2.533	2.459	2.412	2.380	2.339	2.314	2.261	2.206	13
14	4.600	3.739	3.344	3.112	2.958	2.848	2.764	2.699	2.646	2.602	2.565	2.534	2.463	2.388	2.341	2.308	2.266	2.241	2.187	2.131	14
15	4.543	3.682	3.287	3.056	2.901	2.790	2.707	2.641	2.588	2.544	2.507	2.475	2.403	2.328	2.280	2.247	2.204	2.178	2.123	2.066	15
20	4.351	3.493	3.098	2.866	2.711	2.599	2.514	2.447	2.393	2.348	2.310	2.278	2.203	2.124	2.074	2.039	1.994	1.966	1.907	1.843	20

FIGURE 6: The F table only shows values for a 5% one-sided test (or 10% two-sided test, but we don't need that here). We have 30 numerator degrees of freedom and 20 denominator degrees of freedom, and the table says that means we use critical value 2.039.