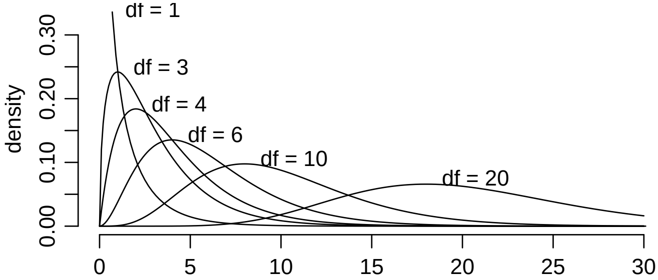
STAT 226 Lecture 1 Supplement

Chi-Squared Distributions and Chi-Squared Tests

Yibi Huang

Chi-Squared (χ^2) Distributions



- One curve for each integer value of **degree of freedom**
- All χ^2-density curves are right-skewed
- χ^2-density curves are only defined on $[0, \infty)$
- As the degrees of freedom ↑, the curves flatten out and move off to the right, and become less skewed (more symmetric)
- Expected value $= df$, SD $= \sqrt{df}$

Relation Between Normal & Chi-Squared Distributions

If there are k independent standard normal variables

$$Z_1, Z_2, \ldots, Z_k \sim N(0, 1),$$

then the sum of their squares

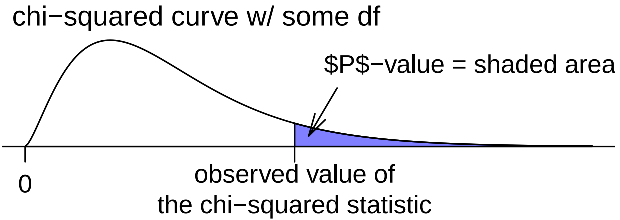
$$Z_1^2 + Z_2^2 + \cdots + Z_k^2$$

has a **Chi-squared** distribution with k **degrees of freedom**.

Chi-Squared Tests in General

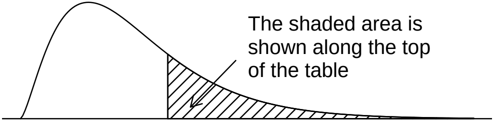
In STAT 226, we will introduce several hypothesis tests that the test-statistic has a **chi-squared** distribution with a certain degrees of freedom.

- The test statistics of such tests are all called **chi-squared (χ^2) statistic**
- For all such tests, large values of chi-squared statistics are evidence against the H_0 and for the H_a, and hence the P-values are *always* the **upper tail** probability below.



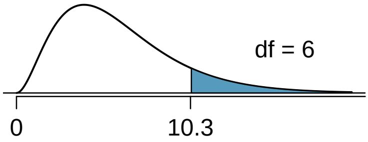
Chi-Square Probability Table

The χ^2-curve, with degrees of freedom shown along the left of the table.



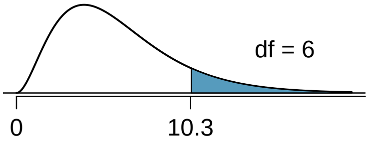
Upper tail	0.3	0.2	0.1	0.05	0.02	0.01	0.005	0.001
df 1	1.07	1.64	2.71	3.84	5.41	6.63	7.88	10.83
2	2.41	3.22	4.61	5.99	7.82	9.21	10.60	13.82
3	3.66	4.64	6.25	7.81	9.84	11.34	12.84	16.27
4	4.88	5.99	7.78	9.49	11.67	13.28	14.86	18.47
5	6.06	7.29	9.24	11.07	13.39	15.09	16.75	20.52
6	7.23	8.56	10.64	12.59	15.03	16.81	18.55	22.46
7	8.38	9.80	12.02	14.07	16.62	18.48	20.28	24.32
8	9.52	11.03	13.36	15.51	18.17	20.09	21.95	26.12
9	10.66	12.24	14.68	16.92	19.68	21.67	23.59	27.88
10	11.78	13.44	15.99	18.31	21.16	23.21	25.19	29.59
11	12.90	14.63	17.28	19.68	22.62	24.72	26.76	31.26
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Ex1. Suppose a chi-squared statistic is 10.3, with df = 6. Find the P-value.



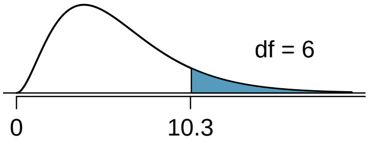
Upper tail		0.3	0.2	0.1	0.05	0.02	0.01	0.005	0.001
df	1	1.07	1.64	2.71	3.84	5.41	6.63	7.88	10.83
	2	2.41	3.22	4.61	5.99	7.82	9.21	10.60	13.82
	3	3.66	4.64	6.25	7.81	9.84	11.34	12.84	16.27
	4	4.88	5.99	7.78	9.49	11.67	13.28	14.86	18.47
	5	6.06	7.29	9.24	11.07	13.39	15.09	16.75	20.52
	6	7.23	8.56	10.64	12.59	15.03	16.81	18.55	22.46
	7	8.38	9.80	12.02	14.07	16.62	18.48	20.28	24.32

Ex1. Suppose a chi-squared statistic is 10.3, with df = 6. Find the P-value.



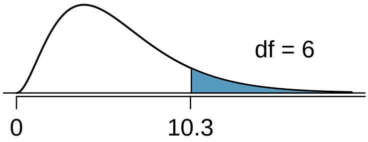
Upper tail		0.3	0.2	0.1	0.05	0.02	0.01	0.005	0.001
df	1	1.07	1.64	2.71	3.84	5.41	6.63	7.88	10.83
	2	2.41	3.22	4.61	5.99	7.82	9.21	10.60	13.82
	3	3.66	4.64	6.25	7.81	9.84	11.34	12.84	16.27
	4	4.88	5.99	7.78	9.49	11.67	13.28	14.86	18.47
	5	6.06	7.29	9.24	11.07	13.39	15.09	16.75	20.52
	6	7.23	8.56	10.64	12.59	15.03	16.81	18.55	22.46
	7	8.38	9.80	12.02	14.07	16.62	18.48	20.28	24.32

Ex1. Suppose a chi-squared statistic is 10.3, with df = 6. Find the P-value.



Upper tail		0.3	*0.2*	*0.1*	0.05	0.02	0.01	0.005	0.001
df	1	1.07	1.64	2.71	3.84	5.41	6.63	7.88	10.83
	2	2.41	3.22	4.61	5.99	7.82	9.21	10.60	13.82
	3	3.66	4.64	6.25	7.81	9.84	11.34	12.84	16.27
	4	4.88	5.99	7.78	9.49	11.67	13.28	14.86	18.47
	5	6.06	7.29	9.24	11.07	13.39	15.09	16.75	20.52
	6	7.23	*8.56*	*10.64*	12.59	15.03	16.81	18.55	22.46
	7	8.38	9.80	12.02	14.07	16.62	18.48	20.28	24.32

Ex1. Suppose a chi-squared statistic is 10.3, with df = 6. Find the P-value.



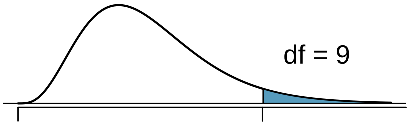
P-value = $P(\chi^2_{df=6} > 10.3)$ is between 0.1 and 0.2

Upper tail		0.3	0.2	0.1	0.05	0.02	0.01	0.005	0.001
df	1	1.07	1.64	2.71	3.84	5.41	6.63	7.88	10.83
	2	2.41	3.22	4.61	5.99	7.82	9.21	10.60	13.82
	3	3.66	4.64	6.25	7.81	9.84	11.34	12.84	16.27
	4	4.88	5.99	7.78	9.49	11.67	13.28	14.86	18.47
	5	6.06	7.29	9.24	11.07	13.39	15.09	16.75	20.52
	6	7.23	8.56	10.64	12.59	15.03	16.81	18.55	22.46
	7	8.38	9.80	12.02	14.07	16.62	18.48	20.28	24.32

The R command `pchisq()` can find the exact area underneath a chi-square curve.

```
pchisq(10.3, df = 6, lower.tail = FALSE)
[1] 0.1126
```

Ex 2. Suppose a χ^2-statistic is 17.56, with df = 9. Find the P-value.

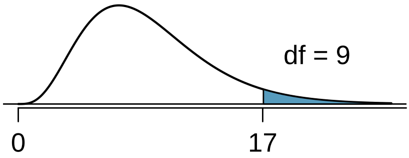


Upper tail		0.3	0.2	0.1	0.05	0.02	0.01	0.005	0.001
df	7	8.38	9.80	12.02	14.07	16.62	18.48	20.28	24.32
	8	9.52	11.03	13.36	15.51	18.17	20.09	21.95	26.12
	9	10.66	12.24	14.68	16.92	19.68	21.67	23.59	27.88
	10	11.78	13.44	15.99	18.31	21.16	23.21	25.19	29.59

```
pchisq(17.56, df = 9, lower.tail = FALSE)
[1] 0.04064
```

P-value = 0.0406.

Ex 2. Suppose a χ^2-statistic is 17.56, with df = 9. Find the P-value.



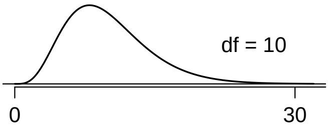
P-value = $P(\chi^2_{df=9} > 17.56)$
is between 0.02 and 0.05

Upper tail	0.3	0.2	0.1	0.05	0.02	0.01	0.005	0.001
df 7	8.38	9.80	12.02	14.07	16.62	18.48	20.28	24.32
8	9.52	11.03	13.36	15.51	18.17	20.09	21.95	26.12
9	10.66	12.24	14.68	16.92	19.68	21.67	23.59	27.88
10	11.78	13.44	15.99	18.31	21.16	23.21	25.19	29.59

```
pchisq(17.56, df = 9, lower.tail = FALSE)
[1] 0.04064
```

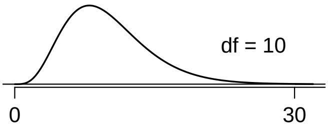
P-value = 0.0406.

Ex 3. Suppose a χ^2-statistic is 30.9, with df = 10. Find the P-value



Upper tail		0.3	0.2	0.1	0.05	0.02	0.01	0.005	0.001
df	8	9.52	11.03	13.36	15.51	18.17	20.09	21.95	26.12
	9	10.66	12.24	14.68	16.92	19.68	21.67	23.59	27.88
	10	11.78	13.44	15.99	18.31	21.16	23.21	25.19	29.59
	11	12.90	14.63	17.28	19.68	22.62	24.72	26.76	31.26

Ex 3. Suppose a χ^2-statistic is 30.9, with df = 10. Find the P-value



P-value = $P(\chi^2_{df=10} > 30.9)$
is less than 0.001

Upper tail		0.3	0.2	0.1	0.05	0.02	0.01	0.005	0.001	→
df	8	9.52	11.03	13.36	15.51	18.17	20.09	21.95	26.12	
	9	10.66	12.24	14.68	16.92	19.68	21.67	23.59	27.88	
	10	11.78	13.44	15.99	18.31	21.16	23.21	25.19	29.59	→
	11	12.90	14.63	17.28	19.68	22.62	24.72	26.76	31.26	

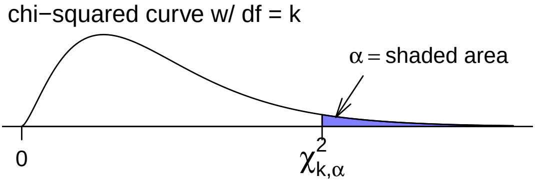
```
pchisq(30.9, df = 10, lower.tail = FALSE)
[1] 0.0006095
```

P-value = 0.0006

Critical Value of a Chi-Squared Test

Just like all other tests, one can also conduct a chi-squared in a critical value approach rather than a P-value approach.

The notation $\chi^2_{\text{df},\alpha}$ represents the value that the upper tail area below a chi-squared curve with df $= k$ to the right of $\chi^2_{k,\alpha}$ is α.



The value $\chi^2_{k,\alpha}$ is called the **critical value** for a chi-squared test with df $= k$ at level α. It's the minimum value for a chi-squared statistic with df $= k$ to be significant at level α.

Finding the Critical Value of a Chi-Squared Test in R

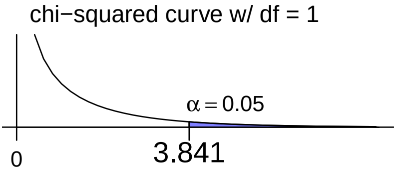
The R command for finding the critical value $\chi^2_{k,\alpha}$ is

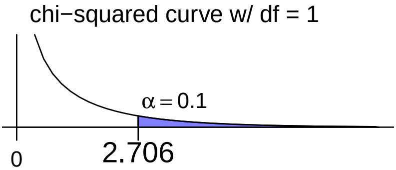
```
qchisq(alpha, df = k, lower.tail=FALSE)
```

Example: If df $= 1$, the chi-squared statistic must be at least

- $\chi^2_{1,0.05} = 3.841$ to be significant at 5% level,and
- $\chi^2_{1,0.1} = 2.706$ to be significant at 10% level.

```
qchisq(0.05, df = 1, lower.tail=FALSE)
[1] 3.841
qchisq(0.1, df = 1, lower.tail=FALSE)
[1] 2.706
```



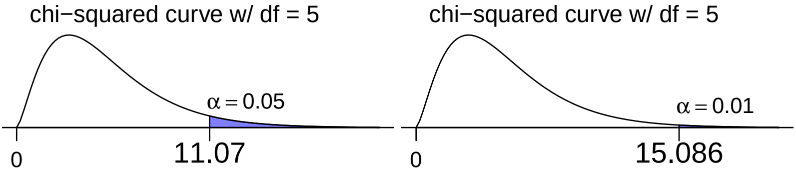


Upper tail		0.3	0.2	0.1	0.05	0.02	0.01	0.005	0.001
df	1	1.07	1.64	2.71	3.84	5.41	6.63	7.88	10.83

Example: A chi-square statistic w/ df = 5 must be at least

- $\chi^2_{5,0.05} = 11.07$ to be significant at 5% level, and
- $\chi^2_{5,0.01} = 15.086$ to be significant at 1% level.

```
qchisq(0.05, df = 5, lower.tail=FALSE)
[1] 11.07
qchisq(0.01, df = 5, lower.tail=FALSE)
[1] 15.09
```



Upper tail	0.3	0.2	0.1	0.05	0.02	0.01	0.005	0.001
df 5	6.06	7.29	9.24	11.07	13.39	15.09	16.75	20.52