comparing independent means

- confidence intervals
- hypothesis tests

Dr. Mine Çetinkaya-Rundel
Duke University

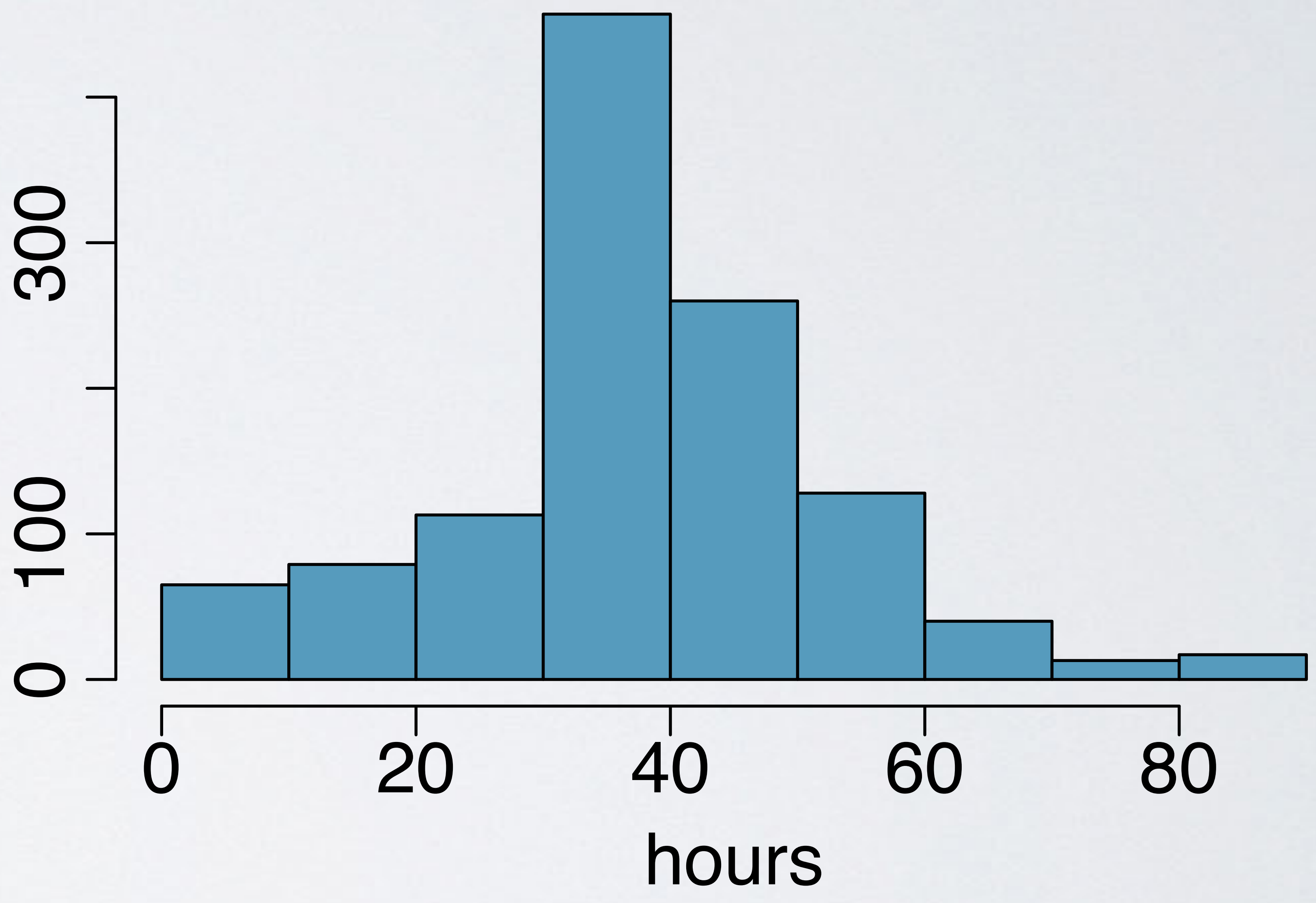
Duke
UNIVERSITY

work hours and educational attainment

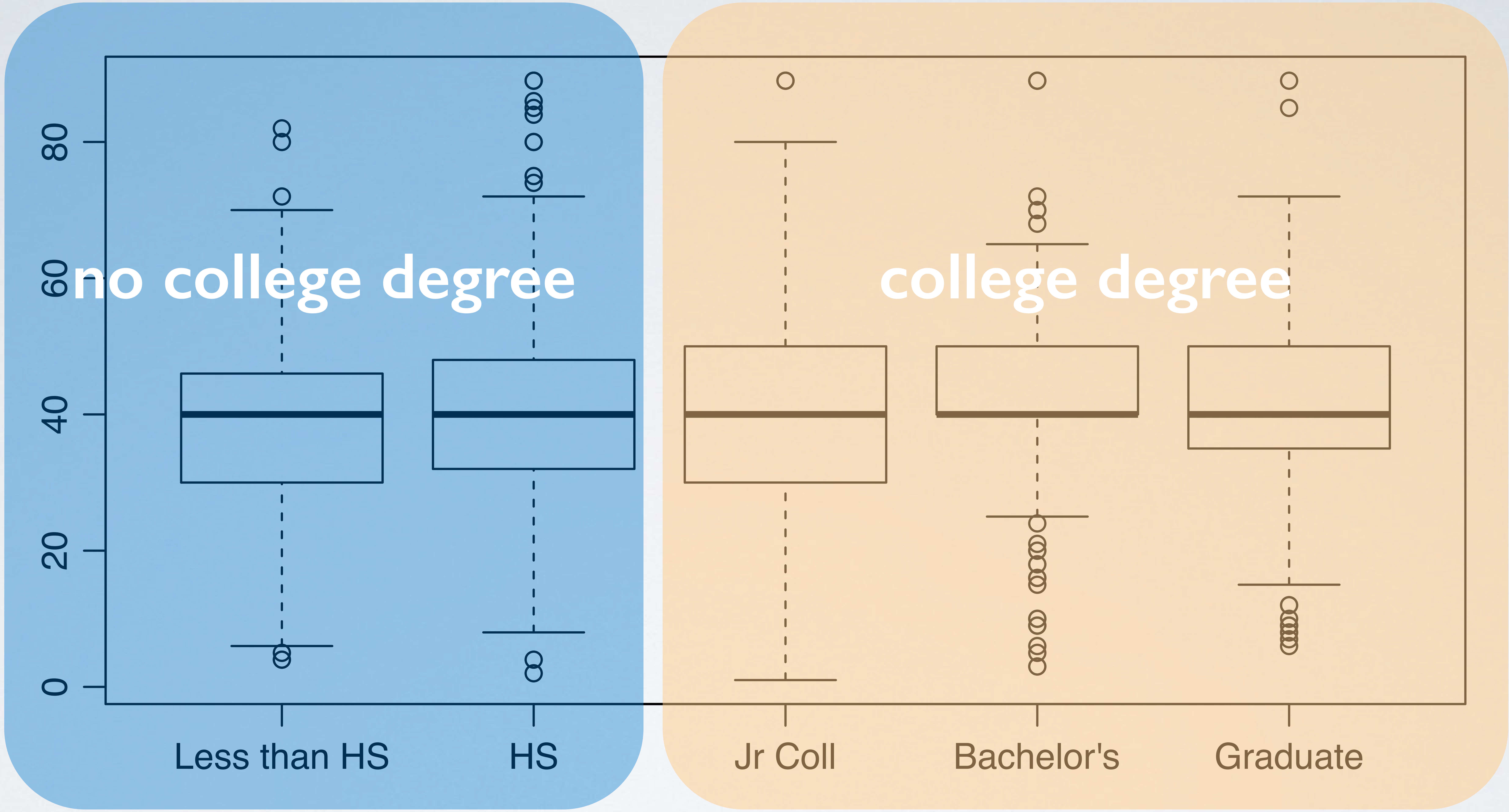
from the 2010 GSS

less than high school	10%
high school	47%
junior college	8%
bachelor's	22%
graduate	13%

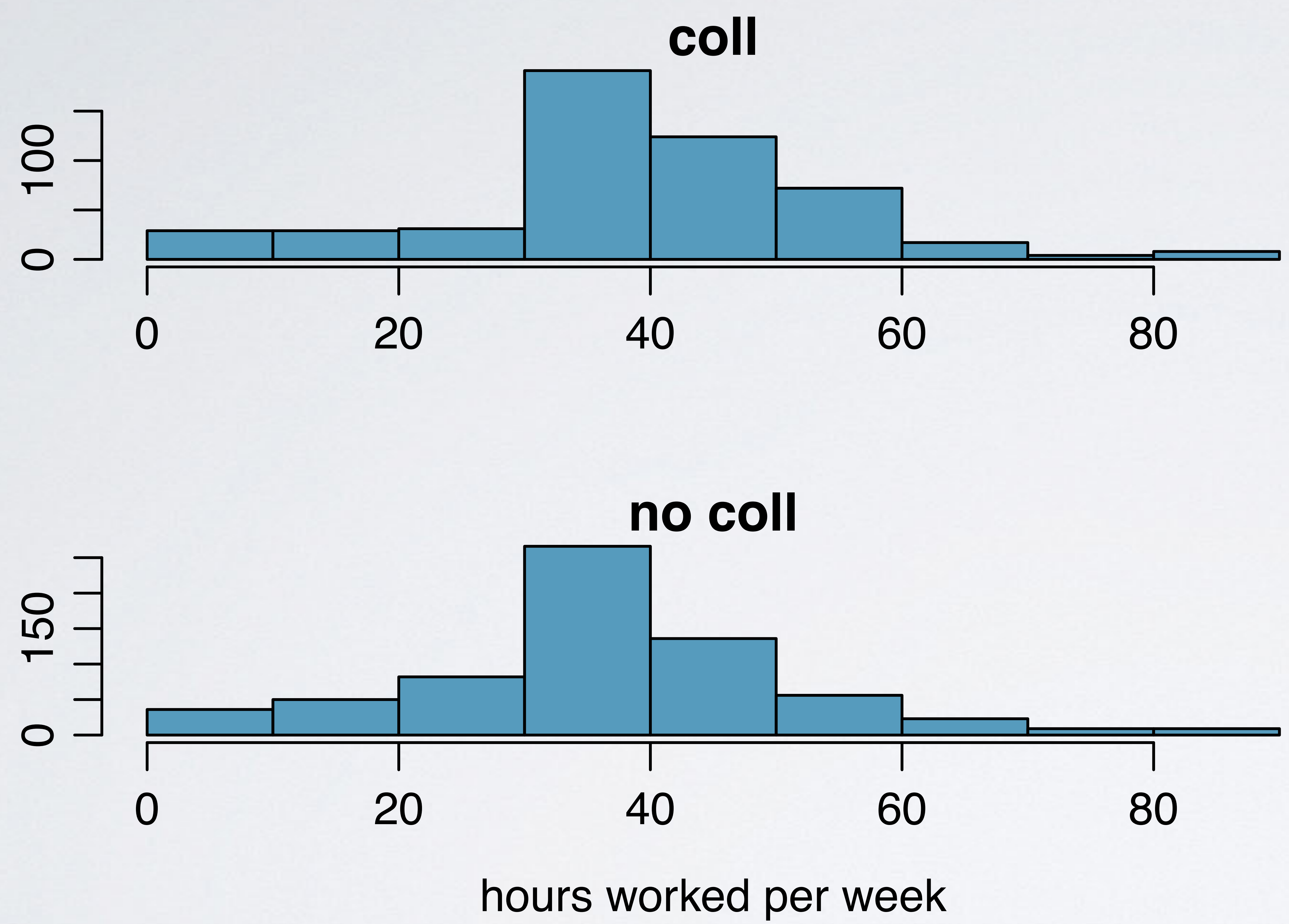
	highest degree	hours
1	bachelor	55
2	bachelor	45
3	junior college	45
...	...	...
1172	high school	40



work hours and educational attainment



another look



	$\bar{x}$	s	n
coll	41.8	15.14	505
no coll	39.4	15.12	667

Estimate how much more (or less) college graduates work, on average, than those without a college degree in the US.

parameter of interest

Average difference between the number of hours worked per week by **all** Americans with a college degree and those without a college degree.

$$\mu_{coll} - \mu_{no\ coll}$$

point estimate

Average difference between the number of hours worked per week by **sampled** Americans with a college degree and those without a college degree.

$$\bar{x}_{coll} - \bar{x}_{no\ coll}$$

estimating the difference between independent means

point estimate ± margin of error

$$(\bar{x}_1 - \bar{x}_2) \pm z^{\star} SE_{\bar{x}_1 - \bar{x}_2}$$

Standard error of difference between two independent means:

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

Conditions for inference for comparing two independent means:

1. ***Independence:***
   - ✓ **within groups:** sampled observations must be independent
     - random sample/assignment
     - if sampling without replacement, $n < 10\%$ of population
   - ✓ **between groups:** the two groups must be independent of each other (non-paired)
2. ***Sample size/skew:*** Each sample size must be at least 30 ($n_1 \geq 30$ and $n_2 \geq 30$), larger if the population distributions are very skewed.

practice

Estimate, using a 95% confidence interval, how much more (or less) college graduates work, on average, than those without a college degree in the US.

	$\bar{x}$	s	n
coll	41.8	15.14	505
no coll	39.4	15.12	667

$$(\bar{x}_{coll} - \bar{x}_{no\ coll}) \pm z^* SE =$$

$$= (41.8 - 39.4) \pm 1.96 \sqrt{\frac{15.14^2}{505} + \frac{15.12^2}{667}}$$

$$= 2.4 \pm 1.96 \times 0.89$$

$$= 2.4 \pm 1.74$$

$$= (0.66, 4.14)$$

makes sense?

testing for a difference between independent means

- null hypothesis: no difference $H_0 : \mu_1 - \mu_2 = 0$
- alternative hypothesis: some difference $H_A : \mu_1 - \mu_2 \neq 0$
- same conditions and SE as the confidence interval

practice

Evaluate whether these data provide convincing evidence that the average number of hours worked by college graduates is different than those without a college degree in the US.

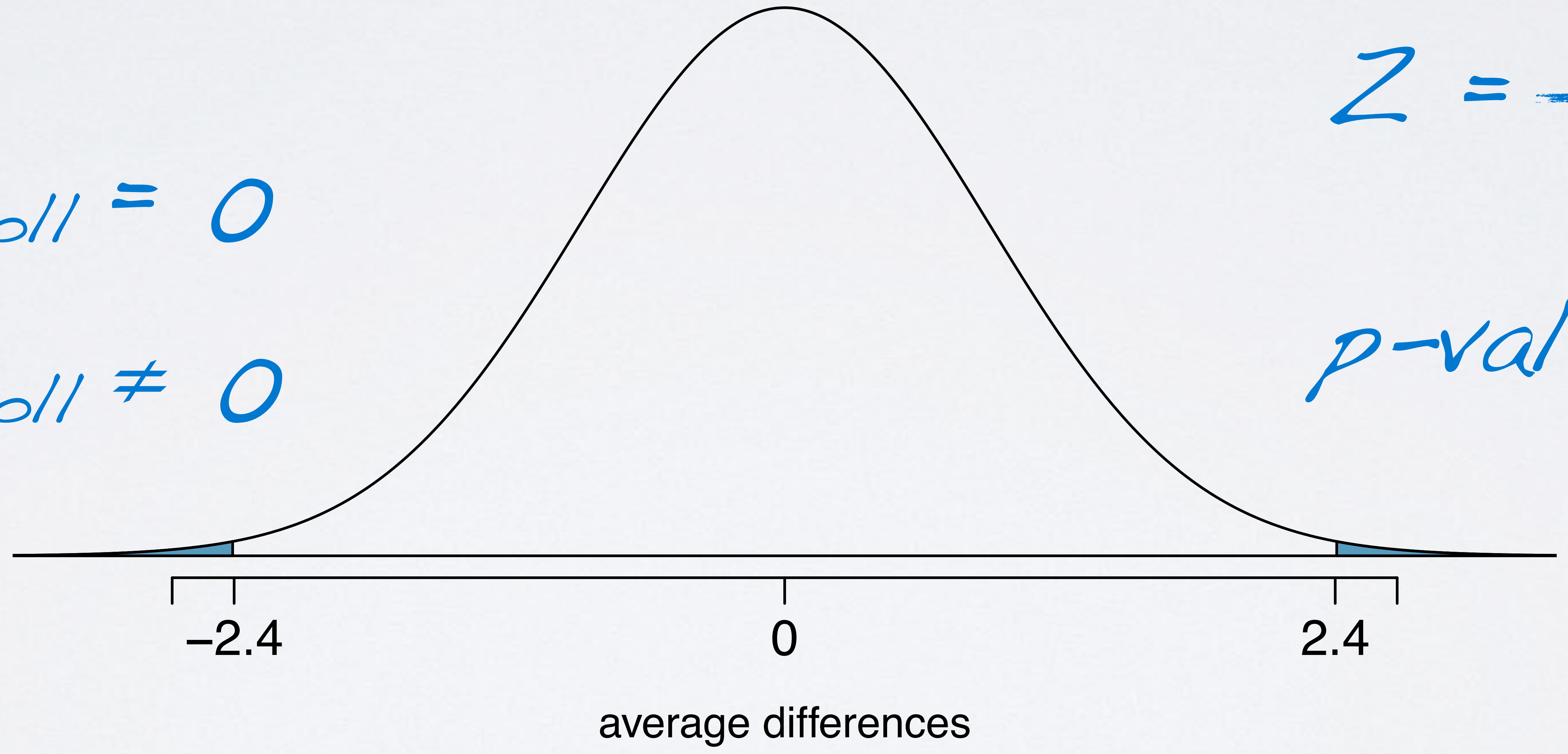
$\bar{x}_{coll} - \bar{x}_{no\ coll} = 2.4$

$SE = 0.89$

$(\bar{X}_{coll} - \bar{X}_{no\ coll}) \sim N(mean = 0, SE = 0.89)$

$H_0: \mu_{coll} - \mu_{no\ coll} = 0$

$H_A: \mu_{coll} - \mu_{no\ coll} \neq 0$



$Z = \dfrac{2.4 - 0}{0.89} = 2.70$

$p\text{-value} = 0.00347 \times 2$

$= 0.00694$

$\approx 0.7\%$

practice

Interpret the p-value (0.7%) in context of the data and the hypotheses.

p-value = P(observed or more extreme statistic | H_0 true)

a difference of 2.4 hours per week

or more, in either direction

there is no difference between the average number of hours worked by those with and without a college degree

If there is no difference between the average number of hours worked by those with and without a college degree, there is a 0.7% chance of obtaining random samples of 505 college and 667 non-college graduates where the average difference between their weekly work hours is at least 2.4 hours.

summary

- independent means
- most often $H_0 : \mu_1 - \mu_2 = 0$
- new formula for SE (everything else the same)

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