section 15: statistical power + ANOVA

stuff to learn today:

1. statistical power (1-β)
2. Welch's t-test
3. Kolmogorov-Smirnov test
4. ANOVA test
5. https://learn.co/tracks/module-2-data-science-career-2-1/statistics-ab-testing-and-linear-regression/section-15-statistical-power-and-anova/welch-s-t-test-lab

statistical power
https://rpsychologist.com/d3/nhst/

- power (1-β) can be defined as:
  - P(rejecting H_0 when it is indeed false)
  - power=1 is a perfect test that always correctly rejects the null hypothesis

- β = P(Type II error)
  = P(failing to reject H_0 when it is indeed false)

- after selecting α, you can typically determine a threshold to maximize the power of a test

Welch's t-test

- Welch's t differs from two-sample Student's t (from Section 14) because it takes into account **different sample sizes** and **unequal variances** between samples

	Welch's t	Student's t
t statistic	$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}}$	$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_p^2(\frac{1}{n_1} + \frac{1}{n_2})}}$
degrees of freedom	$\upsilon \approx \frac{\left(\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}\right)^2}{\frac{s_1^4}{N_1^2 \upsilon_1} + \frac{s_2^4}{N_2^2 \upsilon_2}}$	$n_1 + n_2 - 2$

tests we've seen so far...

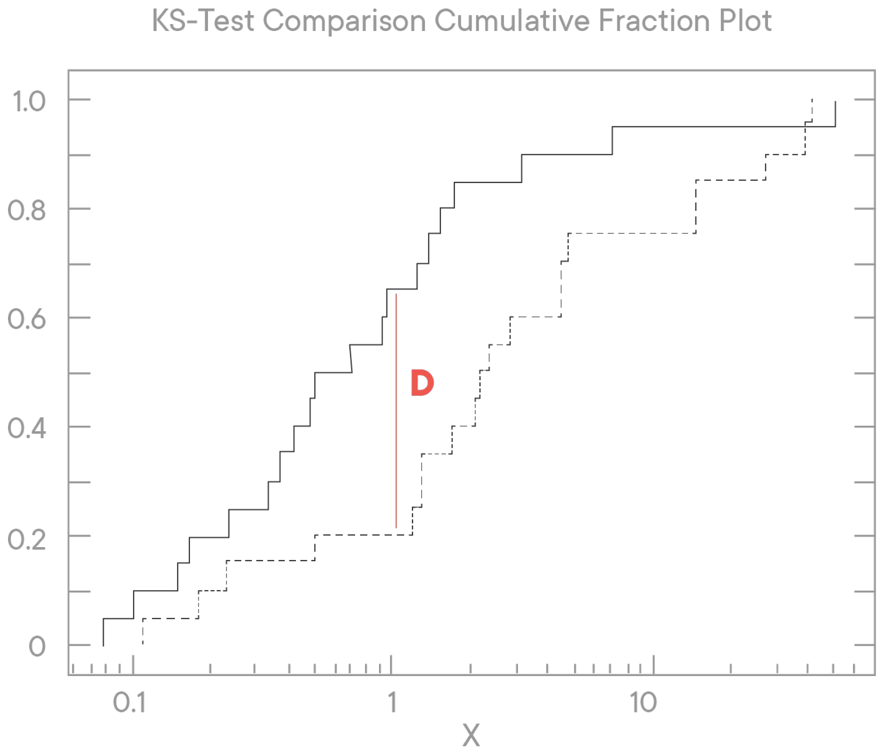
- is one sample is different from the population?
- does tutoring have an effect on IQ of students?
    - 1 sample z-test: when you know your population std
    - 1 sample t-test: when you don't know population std

- are two samples different from each other?
- A/B testing, drug effectiveness testing (placebo vs actual)
    - 2 sample t-test (samples have same size and variance)
    - Welch's t-test (different sizes and variances)

Kolmogorov-Smirnov test

- all the tests we've seen so far have a **normality assumption**

- the 1-sample KS test (K-S Goodness of Fit test) tests for the similarity between a sample and a normal distribution, in terms of the **difference between their cdfs**



Kolmogorov-Smirnov Goodness of Fit Test (1 sample)

- Hypotheses:
  - H_0: No difference between the distribution of the sample and a normal distribution
  - H_A: There is a difference between the distribution of the sample and a normal distribution
- $\alpha = 0.05$
- test statistic = d
- [KS d-table]()

$$d = max(abs[F_0(X) - F_r(X)])$$

- if d > critical d, reject the null hypothesis
  - or compare p value if using scipy

2-sample Kolmogorov-Smirnov test

- Tests whether two samples were drawn from the same population, or two identical populations
- Hypotheses:
  - H_0: No difference between the two distributions
  - H_A: The two distributions are different

- α = 0.05
- test statistic = d
- two-sample d table

$$d = max[abs[F_{n1}(X) - F_{n2}(X)]]$$

- if d > critical d, reject the null hypothesis
  - or compare p value if using scipy

ANOVA test

additional reading

- Compares **group means** with regards to a single variable
- Looking to see if there is a statistically significant difference between groups
- Like doing many t-tests, but gets rid of the *multiple comparisons problem*
  - i.e. a single experiment yields a p-value of 0.03 (3% chance your conclusion to reject H_0 is spurious)
  - So, P(correctly reject H_0) = 0.97
  - P(correctly reject H_0 100 times) = 0.97**100 = 0.0476

- i.e. is there a difference between the mean GPAs of freshmen, sophomores, juniors and seniors?
  - H_0: There is no difference in means
  - H_A: There is a difference in means
  - α = 0.05, test statistic = F

$$F = \frac{\Sigma n_j(\bar{X}_j - \bar{X})^2 / (k-1)}{\Sigma\Sigma(X - \bar{X}_j)^2 / (N-k)}$$

interpreting ANOVA results

data

```
#Your code here
import statsmodels.api as sm
from statsmodels.formula.api import ols

formula = 'len ~ C(supp) + C(dose)'
lm = ols(formula, df).fit()
table = sm.stats.anova_lm(lm, typ=2)
print(table)
```

```
               sum_sq    df          F        PR(>F)
C(supp)     205.350000   1.0  14.016638  4.292793e-04
C(dose)    2426.434333   2.0  82.810935  1.871163e-17
Residual    820.425000  56.0        NaN           NaN
```