Significance tests

- study to address question "Do right-handers miss a target to the right more often than to the left?"
- experiment: each subject throws at target
- variable is X=distance to the target (positive for right, negative for left)
- parameter is mean distance μ for all right-handers
- set up a *null hypothesis* which has form

$$H_0 : \text{parameter} = \text{specific value}$$

- specific value is often, but not always, 0
- for example, set up null hypothesis

$$H_0 : \mu = 0$$

- the *alternative hypothesis* consists of all other reasonable possibilities; here

$$H_a : \mu \neq 0$$

- do experiment with a sample of size 20 and get a sample mean $\bar{x} = 2.4$ cm
- ask "how much evidence does the value we get for $\bar{x}$ provide *against* the null hypothesis?"
- answer by calculating *P-value*: assuming the null hypothesis is true, what is the probability (among all samples of size 20) of getting a sample mean as extreme or more extreme than $\bar{x} = 2.4$ cm (the one we got)

- for our example, suppose we know the distances are normally distributed with a standard deviation of $\sigma = 12.5$ cm
- with this and the null hypothesis $\mu = 0$, have the normal $N(0, 12.5)$ as the distribution for all distances
- for samples of size 20, have

$$\mu_{\bar{x}} = \mu = 0 \qquad \text{and} \qquad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{12.5}{\sqrt{20}} = 2.80$$

  so sample means have distribution $N(0, 2.80)$
- for $\bar{x} = 2.4$ cm, compute

$$z = \frac{2.4 - 0}{2.8} = 0.857$$

- P-value is probability of result as far or further from the mean as the result we got, so

$$\begin{aligned} P(Z < -0.857 \text{ or } Z > 0.857) &= P(Z < -0.857) + P(Z > 0.857) \\ &= 0.1977 + 0.1977 \\ &= 0.3954 \end{aligned}$$

- since this is a high probability, our data provides very weak evidence against the null hypothesis
- conclusion: do not reject the null hypothesis

- given a null hypothesis $\mu = \mu_0$, the *two-sided alternative hypothesis* is $\mu \neq \mu_0$
- in some cases, use a *one-sided alternative hypothesis* of either $\mu < \mu_0$ or $\mu > \mu_0$
  - can use $\mu < \mu_0$ if there is solid reason to think that $\mu > \mu_0$ is not possible
  - can use $\mu > \mu_0$ if there is solid reason to think that $\mu < \mu_0$ is not possible
  - with one-sided alternative, use only one tail in the distribution to determine the P-value
- what counts as a significant P-value?
  - large P-value means measured result is likely if null hypothesis is true so measured result is *weak* evidence *against* null hypothesis
  - small P-value means measured result is unlikely if null hypothesis is true so measured result is *strong* evidence *against* null hypothesis
  - in some contexts, compare P-value to pre-assigned *significance level* α
    - for P-value $> \alpha$, fail to reject H_0
    - for P-value $< \alpha$, reject H_0
- P-value contains more information than decision on "reject or fail to reject"
- best to report P-value along with decision