Lecture Notes Week 8 MSE-213

1 Recap: How to choose a test:

Ask yourself:
- 1 sample vs a number (μ_0) or 2 samples compared to each other?
- Is σ known or N_s very large? $\rightarrow$ z test
- One-sided ("larger", "heavier", "faster", "smaller", ...) or two-sided ("different")?

Note:
- The "significance" $\alpha = 1-P$, where P is the "confidence". Often $P=95\%$, $\alpha=5\%$ or $P=99\%$, $\alpha=1\%$.
- For every cumulative distribution function $CDF(x) = F(x) = \int_{-\infty}^{x} p(t)dt$ we have $\frac{\partial F}{\partial x} \geq 0$ because probabilities can only be ≥ 0.

This means we can compare some P_{TARGET} to a P, or the "score" (i.e. z-value or t-value) directly.

2 Overview of tests so far:

1-Sample Gauss/z-test:

- σ known, or N_s is "large", typically $N_s > 30$ is fine.
- $z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{N_s}}$ some target value
- One-sided: $\Phi(z) \geq P$ or $\Phi(-z) \geq P$ is the criterion
- Two-sided: $2\Phi(|z|) - 1 \geq P$. Find $\Phi(z)$ in table.

Note: $\frac{\partial \Phi(z)}{\partial z} > 0$ so "Bigger z means bigger P"

2-Sample Gauss/z-test

- 2 independent samples (X and Y)
- σ_X and σ_Y known (or N_{sX} and N_{sY} very large)

$$z = \frac{\bar{X} - \bar{Y}}{\sqrt{\sigma_X^2/\sqrt{N_{sX}} + \sigma_Y^2/\sqrt{N_{sY}}}}$$

- Limits: $N_{sX} \rightarrow \infty$: $z = \frac{\bar{X} - \bar{Y}}{\sigma_Y/\sqrt{N_{sY}}}$; $\sigma_X = \sigma_Y = \sigma$, $N_{sY} = N_{sX} = N_s$: $z = \frac{\bar{X} - \bar{Y}}{\sqrt{2}\,\sigma/\sqrt{N_s}}$

1-sample t-test / Student's test

- σ is not known and estimated from the unbiased s (i.e. with $\frac{1}{n-1}$)
- $t = \dfrac{\bar{x} - \bar{y}}{s/\sqrt{N_s}}$ $df = \nu = N_s - 1$ "degrees of freedom"
- Find $CDF(t,\nu) = \tau(t,\nu)$ in table, otherwise same as with the z-test.
- Note: $\dfrac{\partial \tau(t,\nu)}{\partial t} > 0$ and $\dfrac{\partial \tau(t,\nu)}{\partial \nu} > 0$ for $t>0$

2-sample t-test / Student's t-test.

- 2 independent samples x, y.
- We know or assume $\sigma_x = \sigma_y \equiv \sigma$ and estimate σ from the "pooled" unbiased std. dev $s = \sqrt{\dfrac{1}{(N_{sx}-1)+(N_{sy}-1)}\left(\sum_{i=1}^{N_{sx}}(x_i - \bar{x})^2 + \sum_{i=1}^{N_{sy}}(y_i - \bar{y})^2\right)}$
- $t = \dfrac{\bar{x} - \bar{y}}{\sqrt{s^2/N_{sx} + s^2/N_{sy}}}$ (N_{sx} can be $\neq N_{sy}$)
- $df = \nu = (N_{sx} - 1) + (N_{sy} - 1) = N_{TOTAL} - 2$

$\boxed{3}$ 2-sample Welch / Behrens-Fisher test (new)

- 2 independent samples x, y.
- σ_x can be different from σ_y, each is estimated from s_x and s_y resp.

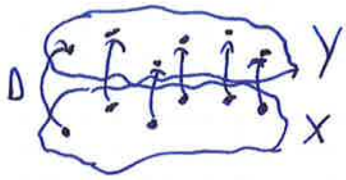
$$t = \frac{\bar{x} - \bar{y}}{\sqrt{s_x^2/N_{sx} + s_y^2/N_{sy}}}$$

$$df = \nu = \underset{\uparrow}{\text{round}}\left\{\left(\frac{s_x^2}{N_{sx}} + \frac{s_y^2}{N_{sy}}\right)^2 \Big/ \left(\frac{(s_x^2/N_{sx})^2}{N_{sx}-1} + \frac{(s_y^2/N_{sy})^2}{N_{sy}-1}\right)\right\}$$

round to nearest integer

- Limits: $N_{sx} \to \infty$ $\nu = N_{sy} - 1$; $\substack{s_x = s_y \\ N_{sx} = N_{sy}}$ $\nu = 2(N_s - 1) = N_{sTOTAL} - 2$

4 The paired/differential 2-Sample test



For example, we measure 25 babies' weight at birth (x) and one week later (y). We can just compare $\bar{x}, s_x$ with $\bar{y}, s_y$. But maybe we are just interested in "how much weight does a baby gain in 1 week"?

Then it makes more sense to look at the pairwise difference first i.e. $d_1 = y_1 - x_1$ etc. (where x_1 and y_1 are the weight of the same baby before and after 1 week).

We then treat D as a random variable, with $\bar{D}$ and σ_D and N_{SD} where N_{SD} is the number of pairs (i.e. 25 here).

D might have a much smaller σ_D than the original σ_x, so it is often a way to get better data.

We then proceed with D as before e.g. a t-test with $t = \frac{\bar{D} - \mu_0}{\sigma_D / \sqrt{N_{SD}}}$

Here often we may have $\mu_0 = 0$ ("did the babies get heavier"), but it could also be $\mu_0 = 100g$ ("did they gain more than 100g").

→ Wherever possible, try to engineer a paired test e.g. measure the same pieces of steel before/after annealing, the velocity of cars before/after entering a tunnel by tracking the same car, the efficiency of the same photodiode in 0 magnetic field or some magnetic field...

5 The χ^2 ("chi-2 / chi-squared) test.

- We have some data – can they be described by some probability distribution function (PDF)?
- Need <u>discrete</u> data / PDF. Otherwise, discretize the data, i.e. chose bins. For distributions such as a normal distribution, we need a bin for $-\infty$ to some number, and some number to $+\infty$.

Example: <u>Null Hypothesis H_0</u>: "We have fair dice" (our friend doesn't cheat)

$$H_0: p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = 1/6 \qquad (\text{Note: } \sum_i p_i = 1)$$

We throw 60 times and count the number for each possible out come. Make a table:

Outcome x_i	⚀	⚁	⚂	⚃	⚄	⚅
Probability p_i	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$
Absolute Frequency n_i	12	9	11	8	10	10
Relative Frequency f_i $f_i = n_i / N_s$	$1/5$	$3/20$	$11/60$	$2/15$	$1/6$	$1/6$

Sample size <u>$N_s = 60$</u>. Number of possible outcomes: <u>$K = 6$</u>

We define a measure that tells us "how much do we deviate from the <u>expectation values</u> of this PDF":

$$\boxed{\chi^2 = \sum_{i=1}^{K} \frac{(n_i - N_s \cdot p_i)^2}{N_s\, p_i} = \sum_{i=1}^{K} N_s \frac{(f_i - p_i)^2}{p_i}}$$

useful for computing

useful for understanding:
- bigger N_s: same relative deviation is a bigger "problem", statistically
- smaller p events get weighted more.

Example:

$$\chi^2 = \frac{(12-10)^2}{10} + \frac{(9-10)^2}{10} + \dots$$

$$= \frac{4}{10} + \frac{1}{10} + \frac{1}{10} + \frac{4}{10} + \frac{0}{10} + \frac{0}{10} = \underline{10/10 = 1}$$

→ The degrees of freedom, given that we impose a PDF with no free parameters are <u>$df = \nu = K - 1$</u> (Example $\nu = 5$)

6 $\chi^2 = 0$ means "we get exactly the expectation values for each possible event". $\chi^2 \geq 0$ always.

$\chi^2 \gg 0$ means "this outcome is unlikely to occur for this probability distribution"

Then we a χ^2 table (with χ^2 and $df = \nu$) to get the corresponding probability. We can <u>reject</u> the null Hypothesis if χ^2 is <u>too large</u>.
For example with $\nu = 5$ (as in our example), if $\chi^2 \geq 11.07$, we can reject the null Hypothesis with 95% confidence (5% significance).

In our example $\chi^2 = 1$ so we <u>cannot</u> claim the die is <u>not</u> fair.

- For the χ^2 test to be a good approximation, the ~~est~~ expectation value in each outcome/bin cannot be to small.

As a rule of thumb, we require: $N_S \cdot \min(p_i) \geq 5$.

outcome w/ the smallest probability

In our example we have $N_S \cdot \frac{1}{6} = 10$.

- This also implies $N_S \gg k$.

- If we estimate some parameters of the PDF from our data, the number of degrees of freedom reduces by that number. For example, if we assume a Gaussian PDF and estimate μ and σ from $\bar{x}$ and s, the degrees of freedom (ν) are reduced by 2.