Summary of NHST for 18.05

Jeremy Orloff and Jonathan Bloom

z-test

- Use: Compare the data mean to an hypothesized mean.
- Data: $x_1, x_2, \ldots, x_n$.
- Assumptions: The data are independent normal samples:
  $x_i \sim N(\mu, \sigma^2)$ where μ is unknown, but σ is known.
- H_0: For a specified μ_0, $\mu = \mu_0$.
- H_A:
  - Two-sided: $\mu \neq \mu_0$
  - one-sided-greater: $\mu > \mu_0$
  - one-sided-less: $\mu < \mu_0$
- Test statistic: $z = \dfrac{\overline{x} - \mu_0}{\sigma/\sqrt{n}}$
- Null distribution: $f(z \,|\, H_0)$ is the pdf of $Z \sim N(0,1)$.
- p-value:
  - Two-sided: $p = P(|Z| > z \,|\, H_0)$ = `2*(1-pnorm(abs(z), 0, 1))`
  - one-sided-greater (right-sided): $p = P(Z > z \,|\, H_0)$ = `1 - pnorm(z, 0, 1)`
  - one-sided-less (left-sided): $p = P(Z < z \,|\, H_0)$ = `pnorm(z, 0, 1)`
- Critical values: z_α has *right*-tail probability α

$$P(z > z_\alpha \mid H_0) = \alpha \quad \Leftrightarrow \quad z_\alpha = \texttt{qnorm}(1-\alpha,\, 0,\, 1).$$

- Rejection regions: let α be the significance.
  - Right-sided rejection region: $[z_\alpha, \infty)$
  - Left-sided rejection region: $(-\infty, z_{1-\alpha}]$
  - Two-sided rejection region: $(-\infty, z_{1-\alpha/2}] \,\cup\, [z_{\alpha/2}, \infty)$

Alternate test statistic

- Test statistic: $\overline{x}$
- Null distribution: $f(\overline{x} \,|\, H_0)$ is the pdf of $\bar{X} \sim \mathrm{N}(\mu_0, \sigma^2/n)$.
- p-value:
  - Two-sided: $p = P(|\bar{X} - \mu_0| > |\overline{x} - \mu_0| \,|\, H_0)$ = `2*(1-pnorm(abs(`$(\overline{x} - \mu_0)$`), 0, `$\sigma/\sqrt{n}$`))`
  - one-sided-greater: $p = P(\bar{X} > \overline{x})$ = `1 - pnorm(`$\overline{x}$`, `μ_0`, `$\sigma/\sqrt{n}$`)`
  - one-sided-less: $p = P(\bar{X} < \overline{x})$ = `pnorm(`$\overline{x}$`, `μ_0`, `$\sigma/\sqrt{n}$`)`
- Critical values: x_α has *right*-tail probability α

$$P(X > x_\alpha \mid H_0) = \alpha \quad \Leftrightarrow \quad x_\alpha = \texttt{qnorm}(1-\alpha,\, \mu_0,\, \sigma/\sqrt{n}).$$

- Rejection regions: let α be the significance.
  - Right-sided rejection region: $[x_\alpha, \infty)$
  - Left-sided rejection region: $(-\infty, x_{1-\alpha}]$
  - Two-sided rejection region: $(-\infty, x_{1-\alpha/2}] \,\cup\, [x_{\alpha/2}, \infty)$

One-sample t-test of the mean

- Use: Compare the data mean to an hypothesized mean.
- Data: $x_1, x_2, \ldots, x_n$.
- Assumptions: The data are independent normal samples:
  $x_i \sim N(\mu, \sigma^2)$ where both μ and σ are unknown.
- H_0: For a specified μ_0, $\mu = \mu_0$
- H_A:

Two-sided:	$\mu \neq \mu_0$
one-sided-greater:	$\mu > \mu_0$
one-sided-less:	$\mu < \mu_0$

- Test statistic: $t = \dfrac{\overline{x} - \mu_0}{s/\sqrt{n}}$,

  where s^2 is the sample variance: $\quad s^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$
- Null distribution: $f(t \,|\, H_0)$ is the pdf of $\ T \sim t(n-1)$.
  (Student t-distribution with $n-1$ degrees of freedom)
- p-value:

Two-sided:	$p = P(\lvert T\rvert > t)$	$=$	`2*(1-pt(abs(t), n-1))`
one-sided-greater:	$p = P(T > t)$	$=$	`1 - pt(t, n-1)`
one-sided-less:	$p = P(T < t)$	$=$	`pt(t, n-1)`

- Critical values: t_α has *right*-tail probability α

$$P(T > t_\alpha \mid H_0) = \alpha \quad \Leftrightarrow \quad t_\alpha = \mathtt{qt}(1-\alpha, n-1).$$

- Rejection regions: let α be the significance.

Right-sided rejection region:	$[t_\alpha, \infty)$
Left-sided rejection region:	$(-\infty, t_{1-\alpha}]$
Two-sided rejection region:	$(-\infty, t_{1-\alpha/2}] \ \cup\ [t_{\alpha/2}, \infty)$

Two-sample t-test for comparing means (assuming equal variance)

- Use: Compare the means from two groups.
- Data: $x_1, x_2, \ldots, x_n$ and $y_1, y_2, \ldots, y_m$.
- Assumptions: Both groups of data are independent normal samples:

$$x_i \sim N(\mu_x, \sigma^2)$$
$$y_j \sim N(\mu_y, \sigma^2)$$

  where both μ_x and μ_y are unknown and possibly different. The variance σ is unknown, but the same for both groups.
- H_0: $\mu_x = \mu_y$
- H_A:

Two-sided:	$\mu_x \neq \mu_y$
one-sided-greater:	$\mu_x > \mu_y$
one-sided-less:	$\mu_x < \mu_y$

- Test statistic: $t = \dfrac{\overline{x} - \overline{y}}{s_P}$,
  where s_x^2 and s_y^2 are the sample variances and s_P^2 is (sometimes called) the pooled sample variance:
  $$s_p^2 = \frac{(n-1)s_x^2 + (m-1)s_y^2}{n+m-2}\left(\frac{1}{n} + \frac{1}{m}\right)$$
- Null distribution: $f(t \mid H_0)$ is the pdf of $T \sim t(n+m-2)$.
  (Student t-distribution with $n+m-2$ degrees of freedom.)
- p-value:

Two-sided:	$p = P(\lvert T\rvert > t)$	$=$	`2*(1-pt(abs(t), n+m-2))`
one-sided-greater:	$p = P(T > t)$	$=$	`1 - pt(t, n+m-2)`
one-sided-less:	$p = P(T < t)$	$=$	`pt(t, n+m-2)`

- Critical values: t_α has *right*-tail probability α
  $$P(t > t_\alpha \mid H_0) = \alpha \quad \Leftrightarrow \quad t_\alpha = \mathtt{qt}(1-\alpha,\, n+m-2).$$
- Rejection regions: let α be the significance.
  Right-sided rejection region: $[t_\alpha, \infty)$
  Left-sided rejection region: $(-\infty, t_{1-\alpha}]$
  Two-sided rejection region: $(-\infty, t_{1-\alpha/2}] \cup [t_{\alpha/2}, \infty)$

Notes: 1. There is a form of the t-test for when the variances are not assumed equal. It is sometimes called Welch's t-test.
2. When the data naturally comes in pairs (x_i, y_i), one uses the *paired two-sample t-test.* For example, in comparing two treatments, each patient receiving treatment 1 might be paired with a patient receiving treatment 2 who is similar in terms of stage of disease, age, sex, etc.

χ^2 test for variance

- Use: Compare the data variance to an hypothesized variance.
- Data: $x_1, x_2, \ldots, x_n$.
- Assumptions: The data are independent normal samples:
  $x_i \sim N(\mu, \sigma^2)$ where both μ and σ are unknown.
- H_0: For a specified σ_0, $\sigma = \sigma_0$
- H_A:
  Two-sided: $\sigma \neq \sigma_0$
  one-sided-greater: $\sigma > \sigma_0$
  one-sided-less: $\sigma < \sigma_0$
- Test statistic: $X^2 = \dfrac{(n-1)s^2}{\sigma_0^2}$, where s^2 is the sample variance: $s^2 = \dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n}(x_i - \overline{x})^2$
- Null distribution: $f(X^2 \mid H_0)$ is the pdf of $\chi^2 \sim \chi^2(n-1)$.
  (Chi-square distribution with $n-1$ degrees of freedom)

- p-value:
  Because the χ^2 distribution is not symmetric around zero the two-sided test is a little awkward to write down. The idea is to look at the X^2 statistic and see if it's in the left or right tail of the distribution. The p-value is twice the probability in that tail.

  An easy check for which tail it's in is: $s^2/\sigma_0^2 > 1$ (right tail) or $s^2/\sigma_0^2 < 1$ (left tail).

$$\text{Two-sided:} \quad p = \begin{cases} 2 * P(\chi^2 > X^2) & \text{if } X^2 \text{ is in the right tail} \\ 2 * P(\chi^2 < X^2) & \text{if } X^2 \text{ is in the left tail} \end{cases}$$
$$= \texttt{2*min(pchisq(}X^2\texttt{,n-1), 1-pchisq(}X^2\texttt{,n-1))}$$

$$\text{one-sided-greater:} \quad p = P(\chi^2 > X^2) = \texttt{1 - pchisq(}X^2\texttt{, n-1)}$$

$$\text{one-sided-less:} \quad p = P(\chi^2 < X^2) = \texttt{pchisq(}X^2\texttt{, n-1)}$$

- Critical values: x_α has *right*-tail probability α

$$P(\chi^2 > x_\alpha \mid H_0) = \alpha \quad \Leftrightarrow \quad x_\alpha = \texttt{qchisq}(1-\alpha, n-1).$$

- Rejection regions: let α be the significance.
  Right-sided rejection region: $[x_\alpha, \infty)$
  Left-sided rejection region: $(-\infty, x_{1-\alpha}]$
  Two-sided rejection region: $(-\infty, x_{1-\alpha/2}] \cup [x_{\alpha/2}, \infty)$

χ^2 test for goodness of fit for categorical data

- Use: Test whether discrete data fits a specific finite probability mass function.
- Data: An observed count O_i in cell i of a table.
- Assumptions: None
- H_0: The data was drawn from a specific discrete distribution.
- H_A: The data was drawn from a different distribution

- Test statistic: The data consists of observed counts O_i for each cell. From the null hypothesis probability table we get a set of expected counts E_i. There are two statistics that we can use:

$$\text{Likelihood ratio statistic } G = 2 * \sum O_i \ln\left(\frac{O_i}{E_i}\right)$$
$$\text{Pearson's chi-square statistic } X^2 = \sum \frac{(O_i - E_i)^2}{E_i}.$$

  It is a theorem that under the null hypthesis $X^2 \approx G$ and both are approximately chi-square. Before computers, X^2 was used because it was easier to compute. Now, it is better to use G although you will still see X^2 used quite often.

- Degrees of freedom df: The number of cell counts that can be freely specified. In the case above, of the n cells $n-1$ can be freely specified and the last must be set to make the correct total. So we have $df = n-1$ degrees of freedom.

  In other chi-square tests there can be more relations between the cell counts os df might be different from $n-1$.

- Rule of thumb: Combine cells until the expected count in each cell is at least 5.
- Null distribution: Assuming H_0, both statistics (approximately) follow a chi-square distribution with *df* degrees of freedom. That is both $f(G \,|\, H_0)$ and $f(X^2 \,|\, H_0)$ have the same pdf as $Y \sim \chi^2(df)$.
- p-value:
  $p \;=\; P(Y > G) \;=\;$ `1 - pchisq(G, df)`
  $p \;=\; P(Y > X^2) \;=\;$ `1 - pchisq(`X^2`, df)`
- Critical values: c_α has *right*-tail probability α

$$P(Y > c_\alpha \mid H_0) = \alpha \quad \Leftrightarrow \quad c_\alpha = \texttt{qchisq}(1-\alpha,\, df).$$

- Rejection regions: let α be the significance.
  We expect X^2 to be small if the fit of the data to the hypothesized distribution is good. So we only use a right-sided rejection region: $[c_\alpha,\, \infty)$.

One-way ANOVA (F-test for equal means)

- Use: Compare the data means from n groups with m data points in each group.
- Data:

$$\begin{array}{cccc} x_{1,1}, & x_{1,2}, & \ldots, & x_{1,m} \\ x_{2,1}, & x_{2,2}, & \ldots, & x_{2,m} \\ & & \ldots & \\ x_{n,1}, & x_{n,2}, & \ldots, & x_{n,m} \end{array}$$

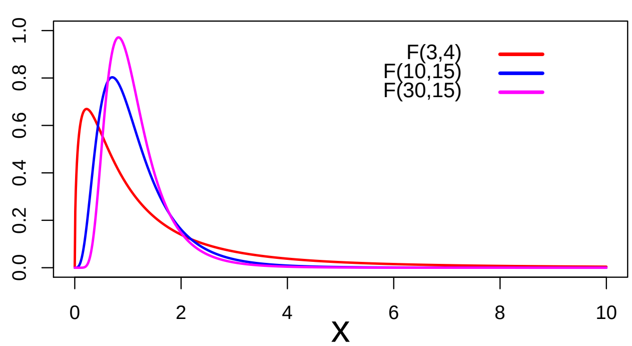
- Assumptions: Data for each group is an independent normal sample drawn from distributions with (possibly) different means but the same variance:

$$\begin{array}{ll} x_{1,j} & \sim N(\mu_1, \sigma^2) \\ x_{2,j} & \sim N(\mu_2, \sigma^2) \\ & \ldots \\ x_{n,j} & \sim N(\mu_n, \sigma^2) \end{array}$$

  The group means μ_i are unknown and possibly different. The variance σ is unknown, but the same for all groups.
- H_0: All the means are identical $\mu_1 = \mu_2 = \ldots = \mu_n$.
- H_A: Not all the means are the same.
- Test statistic: $w = \frac{\text{MS}_B}{\text{MS}_W}$, where

$$
\begin{aligned}
\bar{x}_i &= \text{mean of group } i \\
&= \frac{x_{i,1} + x_{i,2} + \ldots + x_{i,m}}{m}. \\
\overline{x} &= \text{grand mean of all the data.} \\
s_i^2 &= \text{sample variance of group } i \\
&= \frac{1}{m-1} \sum_{j=1}^{m} (x_{i,j} - \bar{x}_i)^2. \\
\text{MS}_B &= \text{between group variance} \\
&= m \times \text{sample variance of group means} \\
&= \frac{m}{n-1} \sum_{i=1}^{n} (\bar{x}_i - \overline{x})^2. \\
\text{MS}_W &= \text{average within group variance} \\
&= \text{sample mean of } s_1^2, \ldots, s_n^2 \\
&= \frac{s_1^2 + s_2^2 + \ldots + s_n^2}{n}
\end{aligned}
$$

- Idea: If the μ_i are all equal, this ratio should be near 1. If they are not equal then MS_B should be larger while MS_W should remain about the same, so w should be larger. We won't give a proof of this.
- Null distribution: $f(w \,|\, H_0)$ is the pdf of $W \sim F(n-1, n(m-1))$.
  This is the F-distribution with $(n-1)$ and $n(m-1)$ degrees of freedom. Several F-distributions are plotted below.
- p-value: $p = P(W > w) =$ `1- pf(w, n-1, n*(m-1)))`



Notes: 1. ANOVA tests whether all the means are the same. It does not test whether some subset of the means are the same.
2. There is a test where the variances are not assumed equal.
3. There is a test where the groups don't all have the same number of samples.

F-test for equal variances

- Use: Compare the vaiances from two groups.
- Data: $x_1, x_2, \ldots, x_n$ and $y_1, y_2, \ldots, y_m$.
- Assumptions: Both groups of data are independent normal samples:

$$
\begin{aligned}
x_i &\sim N(\mu_x, \sigma_x^2) \\
y_j &\sim N(\mu_y, \sigma_y^2)
\end{aligned}
$$

where μ_x, μ_y, σ_x and σ_y are all unknown.

- H_0: $\sigma_x = \sigma_y$
- H_A:
  - Two-sided: $\sigma_x \neq \sigma_y$
  - one-sided-greater: $\sigma_x > \sigma_y$
  - one-sided-less: $\sigma_x < \sigma_y$
- Test statistic: $w = \dfrac{s_x^2}{s_y^2}$,

  where s_x^2 and s_y^2 are the sample variances of the data.
- Null distribution: $f(w \mid H_0)$ is the pdf of $W \sim F(n-1, m-1)$.
  (F-distribution with $n-1$ and $m-1$ degrees of freedom.)
- p-value:

Two-sided:	p	=	`2*min(pf(w,n-1,m-1), 1-pf(w, n-1,m-1))`
one-sided-greater:	$p = P(W > w)$	=	`1 - pf(w, n-1, m-1)`
one-sided-less:	$p = P(W < w)$	=	`pf(w, n-1, m-1)`

- Critical values: w_α has *right*-tail probability α

$$P(W > w_\alpha \mid H_0) = \alpha \quad \Leftrightarrow \quad w_\alpha = \texttt{qf}(1-\alpha, n-1, m-1).$$