Popular Discrete and Continuous Distributions

Bayes' Theorem

$$P(A_j \mid B) = \frac{P(B \mid A_j)P(A_j)}{P(B)} = \frac{P(B \mid A_j)P(A_j)}{\sum_i P(B \mid A_i)P(A_i)}$$

probably besides Bayes' Theorem literally everything else on this page isn't necessary because we do it in R anyway.

Binomial(n,p)

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0,1,2,\ldots$$

$$E[X] = np \quad Var[X] = np(1-p)$$

also remember this for conditional probability:

Conditional Probability

$$Pr(A|B) = \frac{Pr(A \cap B)}{Pr(B)}$$

Poisson(λ)

$$f(x) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0,1,2,\ldots$$

$$E[X] = \lambda \quad Var[X] = \lambda$$

for continuous random variables remember this (idk why this isn't on here bruh):

General formula for continuous rv:

$$P(a < X \le b) = \int_a^b f(x)\,dx$$

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(w)\,dw$$

expected value (aka the mean)

$$E[X] = \int_{-\infty}^{\infty} x \cdot f(x)\,dx \quad \text{and} \quad E[u(X)] = \int_{-\infty}^{\infty} u(x) f(x)\,dx$$

$$V[X] = E[X^2] - (E[X])^2$$

variance
(aka standard deviation squared)

NegativeBinomial(r,p)

$$f(x) = \binom{x+r-1}{r-1} p^r (1-p)^x, \quad x = 0,1,2,\ldots$$

$$E[X] = \frac{r(1-p)}{p} \quad Var[X] = \frac{r(1-p)}{p^2}$$

Geometric(p)

$$f(x) = p(1-p)^x, \quad x = 0,1,2,\ldots$$

$$E[X] = \frac{1-p}{p} \quad Var[X] = \frac{1-p}{p^2} \quad M_X(t) = \frac{p}{1-(1-p)e^t}, \; t \in \mathbb{R}, (1-p)e^t < 1$$

Discrete Random Variables:
if they give you a table
with x values and P(x) values
then:

make x and the px into two datasets
then do: sum(x*px)

for variance: sum((x^2)*px) - (sum(x*px))^2

Hypergeometric

$$f(x) = \frac{\binom{N_1}{x}\binom{N_2}{n-x}}{\binom{N_1+N_2}{n}}, \quad x = 0,1,2,\ldots,n$$

Gamma(α, β)

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta}, \quad 0 \le x < \infty \qquad \text{Note: } \Gamma(r) = (r-1)!, \text{ for all } r = 1,2,3,\ldots$$

$E[X] = \alpha\beta \quad Var[X] = \alpha\beta^2$

Normal(μ, σ)

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}$$

$E[X] = \mu \quad Var[X] = \sigma^2$

Exponential (λ)

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \ge 0 \\ 0 & otherwise \end{cases}$$

$E[X] = 1/\lambda$

$Var[X] = 1/\lambda^2$

other than looking at the means and variances for each of these, all the other stuff on this page isn't that helpful.

Remember this:

> if its a percentage in the problem, use binomial, if its poisson they'll tell you its poisson otherwise its hypergeometric
> -- Nick

(they will also almost always tell you if its normal distribution or not)

R "base" commands for distributions:

"_" filled in with d, p or q

```
_binom( )
_exp( )
_gamma( )
_hyper( )
_nbinom( )
_norm( )
_pois( )
_weibull( )
_t( )
_f( )
_tukey( )
```

some notes about normal distribution:

pnorm(x, mean, sd) $P(x \le X)$

1 - pnorm(x, mean, sd) $P(x \ge X)$

pnorm(b, mean, sd) - pnorm(a, mean, sd)

$P(a \le x \le b)$

if they give you a sample size as well
do the following: pnorm(x, mean, (sd/sqrt(n)))
(same applies if its greater or between a range)

qnorm() gives you the inverse
(you'll use this to get critical values for z tests later)

Formulas for joint distributions:

$$P(Y=y\mid X=x)=\frac{P(X=x,Y=y)}{P(X=x)}=\frac{p(x,y)}{p_X(x)}$$

$$P(X=x\mid Y=y)=\frac{P(X=x,Y=y)}{P(Y=y)}=\frac{p(x,y)}{p_Y(y)}$$

$$\mu_X=E[X]=\sum_{(x,y)\in{}^{\circ}} x\,p(x,y)=\sum_{x} x\,p_X(x)$$

$$\mu_Y=E[Y]=\sum_{(x,y)\in{}^{\circ}} y\,p(x,y)=\sum_{y} y\,p_Y(y)$$

$$E\left[g(X,Y)\right]=\sum_{(x,y)\in{}^{\circ}} g(x,y)\,p(x,y)$$

$$\operatorname{cov}(X,Y)=E[XY]-E[X]E[Y]$$

$$\rho=\frac{\operatorname{cov}(X,Y)}{\sigma_X\sigma_Y}$$

Hypothesis tests and confidence intervals:

Test	Null Hypothesis	Test Statistic
One-sample z-test for means	$\mu=\mu_o$	$z=\dfrac{\bar{x}-\mu_o}{\frac{\sigma}{\sqrt{n}}}$

use when you know population standard deviation

for p-value:

left tailed: pnorm(z)

right tailed: 1 - pnorm(z)

two tailed:
if z is positive: 2*(1-pnorm(z))
if z is negative: 2*pnorm(z)

Test	Null Hypothesis	Test Statistic
One-sample t-test for means	$\mu=\mu_o$	$t=\dfrac{\bar{x}-\mu_o}{\frac{s}{\sqrt{n}}}$; df = n–1

use when you only have sample standard deviation

for p-value:

left tailed: pt(t, n-1)

right tailed: 1 - pt(t, n-1)

two tailed:
if t is positive: 2*(1-pt(t, n-1))
if t is negative: 2*pt(t, n-1)

Test	Null Hypothesis	Test Statistic
Matched Pairs t-test	$\mu_D=\mu_{D_0}$	$t=\dfrac{\bar{w}-\mu_{D_0}}{s/\sqrt{n}}$; df = n – 1

t.test(before, after, alternative = "greater", paired = T)

this is if its right tailed

if left tailed do: alternative = "less"
if two tailed do: alternative = "two.sided"

the t.test function will
give you the p value
(if doing it manually just do pt)

Test	Null Hypothesis	Test Statistic
One-sample z-test for proportions	$p=p_o$	$z=\dfrac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}}$

use pnorm in the same manner stated at the beginning

Two-sample t-test for means $\mu_1 - \mu_2 = 0$ or $\mu_1 = \mu_2$ $t = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$; df=$v$

t.test(dataSet1, dataSet2, alternative = "greater")

this is if its right tailed

if left tailed do: alternative = "less"

if two tailed do: alternative = "two.sided"

$$v = \frac{\left(\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}\right)^2}{\dfrac{(s_1^2/n_1)^2}{n_1 - 1} + \dfrac{(s_2^2/n_2)^2}{n_2 - 1}}$$

p-value given by the function

Two-sample z-test for means $\mu_1 - \mu_2 = 0$ or $\mu_1 = \mu_2$ $z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$ or $z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$

Two-sample z-test for proportions $p_1 - p_2 = 0$ or $p_1 = p_2$ $z = \dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$

use pnorm for p-value in same manner as stated earlier

χ^2 - Goodness of Fit __ is same as _ $\chi^2 = \sum \dfrac{(O-E)^2}{E}$

chisq.test(dataSet)

if also given a set of proportions then do: chisq.test(dataSet, p = proportionsDataSet)

<u>Confidence Intervals</u>

One-sample z-test: $\bar{x} \pm z^* \dfrac{\sigma}{\sqrt{n}}$ xBar + c(-1,1) * qnorm((1+C)/2) * (sigma/sqrt(n))

One-sample t-test: $\bar{x} \pm t^* \dfrac{s}{\sqrt{n}}$ xBar + c(-1,1) * qt((1+C)/2, n-1) * (sd/sqrt(n))

One-proportion z-test: $\hat{p} \pm z^* \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$ pHat + c(-1,1) * qnorm((1+C)/2) * sqrt((pHat*(1-pHat))/n)

Two-sample t-test: $(\bar{x}_1 - \bar{x}_2) \pm t^* \sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}$ (xBar1 - xBar2) + c(-1,1) * qt((1+C)/2, df) * sqrt((s1^2/n1) + (s2^2/n2))
df is the smallest n minus 1

if you have matched pairs you could also just do: t.test(before, after, conf.level = decimalValue, paired = T)

Two-sample z-test: $(\bar{x}_1 - \bar{x}_2) \pm z^* \sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}$ do the same manual command as above but use qnorm instead
(tho I'll be honest I don't think we're gonna get a two sample z conf. interval)

Two-proportion z-test: $(\hat{p}_1 - \hat{p}_2) \pm z^* \sqrt{\dfrac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \dfrac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$

(pHat1 - pHat2) + c(-1,1) * qnorm((1+C)/2) * sqrt(((pHat1*(1-pHat1))/n1) + ((pHat2*(1-pHat2))/n2)))
(you probs don't need as many parenthesis here, I just wanted to make sure I was separating them correctly)

Confidence interval for σ^2: $\left[\dfrac{(n-1)S^2}{\chi^2_{\alpha/2}}, \dfrac{(n-1)S^2}{\chi^2_{1-\alpha/2}}\right]$

χ^2 has $n-1$ degrees of freedom.

lower bound: ((n-1)*s^2)/qchisq((1+C)/2, n-1, lower.tail=F)
upper bound: ((n-1)*s^2)/qchisq((1+C)/2, n-1, lower.tail=T)

if it asks for confidence interval of standard deviation
take the square root of both of those commands

Slope of regression line: $(\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x)$: $\hat{\beta}_1 \pm t_{\alpha/2} SE_{\hat{\beta}_1}$

if you already have a lm variable setup in R
you can just do: confint(name.lm, level = 0.95)
(replace 0.95 with whatever conf level)

However if you haven't been given the linear model you can do this command to get the confidence interval:

b1 + c(-1,1) * qt((1+C)/2, n-2) * standardError

b1 is the thing you multiply to x in the equation
that and standardError will be given to you.
Note: for this one degrees of freedom is n-2

Tukey's w: $w = Q_{\alpha, M, N-M} \sqrt{\dfrac{MS(resid)}{N/M}}$ (outside of one example in the slides, we've basically never used this so I'm not gonna bother for this one lol)