Ch 1 – 3 μ or $\bar{x} = \frac{\Sigma x_i}{n}$

range = largest – smallest value

unusual if $p < 0.05$

median : if n = odd , median = middle number

if n = even , median = $\frac{\text{2 middle number}}{2}$

midpoint = $\frac{\text{next lower} - \text{lower}}{2}$

$$\sigma = \sqrt{\frac{\Sigma (x_i - \mu)^2}{n}} = \frac{\Sigma(\text{midpoint} - \mu)^2 \cdot \text{frequency}}{n}$$

$$s = \sqrt{\frac{\Sigma (x_i - \bar{x})^2}{n-1}} = \frac{\Sigma(\text{midpoint} - \bar{x})^2 \cdot \text{frequency}}{n}$$

68% of data within 1 standard deviation are between $\mu - \sigma$ and $\mu + \sigma$

95% 2

99% or almost 3

If Q_1 , $L = 0.25n$, if L = whole number , $Q = \frac{L + (L+1)}{2}$

Q_3 , $L = 0.75n$ $\neq$, $Q = \text{ceil}(L)$

For p^{th} percentile , $L = (\frac{p}{100})n$

---

percentile = $100 \left(\frac{(\#\text{ of values less than } x) + 0.5}{n}\right)$, alway round up

---

interquartile range = IQR = $Q_3 - Q_1$

lower outlier boundary = $Q_1 - 1.5(IQR)$

upper outlier boundary = $Q_3 + 1.5(IQR)$

Ch 5 – 7 For any two event A and B

if A and B are mutually exclusive : $P(A \text{ or } B) = P(A) + P(B)$

else : $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$

Complement : $P(A^c) = 1 - P(A)$; $P(A) + P(A^c) = 1$; $P(A) = 1 - P(A^c)$

Probability of event B given event A = $P(B \mid A) = \frac{P(A \text{ and } B)}{P(A)}$

$P(A \text{ and } B) = P(A) \cdot P(B \mid A) = P(A) \cdot P(A \mid B)$

$P(A \text{ and } B) = 0$, if A and B are mutually exclusive

mutually exclusive : impossible for both events to occur altogether

---

independent : one does not affect that the other events to occur

if A and B are independent events , then $P(A \text{ and } B) = P(A) \cdot P(B)$

when sampling with replacement, the draws are independent

$P(\text{at least one}) = 1 - (1-p)^n = 1 - P(\text{non occurred})$

---

mean of random variable = expected value = $\mu_x = E(x) = \Sigma(x \cdot p(x))$

standard deviation of random variable = $\sigma_x = \sqrt{\Sigma[(x - \mu_x)^2 \cdot p(x)]}$

binompdf : $nC_x p^x (1-p)^{n-x} = p(x = k) = \text{binompdf}(n, p, x)$

binomcdf : $p(x \leq k) = 1 - p(x > k) = \text{binomcdf}(n, p, x)$

mean of binomial random variable : $\mu_x = n \cdot p$

standard deviation of binomial random variable : $\sigma_x = \sqrt{n \cdot p(1-p)}$

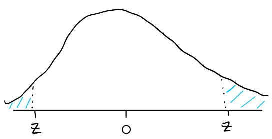
---

z – score : $z = \frac{x - \mu}{\sigma}$ = invNorm (area , μ , σ , tail)

Area of a given z-score : normalcdf (lower, upper, μ , σ)

x corresponds to a given z – score : $x = \mu + z \cdot \sigma$

$z = n^{th}$ percentile from left



Central limit theorem :

$\mu_{\bar{x}} = \mu$, $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

$\hat{p} = \frac{x}{n}$

mean of $\hat{p}$: $\mu_{\hat{p}} = p$

standard deviation of $\hat{p}$: $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$

Ch 8,9,11

point estimate : $\bar{x}$ or $\hat{p}$, critical value : z-score, standard error : $\frac{\theta}{\sqrt{n}}$, $\frac{\alpha}{2} = \frac{1-\text{C-level}}{2}$,

Sample size : $n = \left(\frac{z \cdot \theta}{m}\right)^2$, margin of error = (critical value)·(standard error)

C interval : $\bar{x} - m < \mu < \bar{x} + m$

$\hat{p} - m < p < \hat{p} + m$

Construct Confidence interval

known θ

z-method : (Z interval)

$z_{\alpha/2}$ = invNorm(C-level, 0, 1, center)

$m = z_{\alpha/2} \cdot \frac{\theta}{\sqrt{n}}$

unknown θ

T-method : (T interval)

$t_{\alpha/2}$ = invT$\left(1-\frac{\alpha}{2}, n-1\right)$

$m = t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$

C-interval for proportion : (1 prop Z Int)

$\hat{p} = \frac{x}{n}$

$z_{\alpha/2}$ = invNorm(C-level, 0, 1, center)

$m = z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

$n = \hat{p}(1-\hat{p})\left(\frac{z_{\alpha/2}}{m}\right)^2$ or $0.25\left(\frac{z_{\alpha/2}}{m}\right)^2$

critical values using chi-square :

$\chi^2_{1-\frac{\alpha}{2}} = \left(1-\frac{\alpha}{2}, df\right)$

$\chi^2_{\frac{\alpha}{2}} = \left(\frac{\alpha}{2}, df\right)$

C-interval for population θ is

$$\sqrt{\frac{(n-1)s^2}{\chi^2_{\alpha/2}}} < \theta < \sqrt{\frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}}$$

$$\frac{\alpha}{2} = \frac{1-(\text{c-level})}{2}$$

null hypothesis : $H_0 : \mu = \mu_0$

alternate hypothesis : $H_1 : \mu < \mu_0,\ \mu > \mu_0,\ \mu \neq \mu_0$

level of significance : $\alpha = 0.05$ (if not mentioned)

Type I error : reject H_0 when H_0 is true ($p > \alpha$)

Type II error : do not reject H_0 when H_0 is false ($p < \alpha$)

If $p < \alpha$, reject H_0. Enough evidence

If $p > \alpha$, do not reject H_0. Not enough evidence

Smaller p is, stronger against H_0

Hypothesis test

Known θ

z-test :

Test statistic : $z = \frac{\bar{x} - \mu_0}{\left(\frac{\theta}{\sqrt{n}}\right)}$

p-value :

p = normalcdf($-\infty$, z, 0, 1) left-tailed

p = normalcdf(z, ∞, 0, 1) right-tailed

p = 2·normalcdf(z, ∞, 0, 1) two-tailed

unknown θ

T Test :

Test Statistic : $t = \frac{\bar{x} - \mu_0}{\left(\frac{s}{\sqrt{n}}\right)}$

p-value :

p = tcdf($-\infty$, t, $n-1$) left-tailed

p = tcdf(t, ∞, $n-1$) right-tailed

2·tcdf(t, ∞, $n-1$) two-tailed

Hypothesis test for proportion

$H_0 : p = p_0$

$H_1 : p < p_0,\ p > p_0,\ p \neq p_0$

1 prop Z Test

Test statistic : $z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$

p-value = normalcdf($-\infty$, z, 0, 1) left-tailed

normalcdf(z, ∞, 0, 1) right-tailed

2·normalcdf(z, ∞, 0, 1) two-tailed

$H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 < \mu_2,\ \mu_1 > \mu_2,\ \mu_1 \neq \mu_2$

standard error of $\bar{x}_1 - \bar{x}_2 = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

degree of freedom : smaller of $n_1 - 1$ and $n_2 - 1$

Two means : Independent samples

2-SampTTest :

Test statistic : $t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$

p-value : p = tcdf($-\infty$, z, df) left-tailed

tcdf(z, ∞, df) right-tailed

2·tcdf(z, ∞, df) two-tailed

$H_0 : p_1 = p_2$

$H_1 : p_1 < p_2,\ p_1 > p_2,\ p_1 \neq p_2$

mean = $p_1 - p_2$, standard deviation = $\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$

$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$, standard error = $\sqrt{\frac{\hat{p}(1-\hat{p})}{n_1} + \frac{\hat{p}(1-\hat{p})}{n_2}} = \sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$

Two proportions

2-propZTest :

Test statistic : $z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$

p-value : p = normalcdf($-\infty$, z, ∞, 0, 1) left-tailed / normalcdf(z, ∞, 0, 1) right-tailed

matched pairs : dependent samples

$d = \bar{x}_1 - \bar{x}_2$

$\bar{d}$ = mean of d

$H_0 : \mu_d = 0$

$H_1 : \mu_d < 0,\ \mu_d > 0,\ \mu_d \neq 0$

Two means : paired samples

T Test :

Test statistic : $t = \frac{\bar{d} - \mu_0}{\left(\frac{s_d}{\sqrt{n_d}}\right)}$ always 0

p-value : p = tcdf(z, ∞, $n_d - 1$)

Assumptions : SRS and $n > 30$ or normally distributed

Assumptions for proportion : SRS, population $\geq 20 \cdot n$, categories = 2

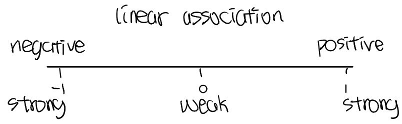
and each categories > 10

Ch 4 correlation coefficient : $r = \frac{1}{n-1} \Sigma \left(\frac{x-\bar{x}}{S_x}\right)\left(\frac{y-\bar{y}}{S_y}\right)$

To find $ax+b$: STAT → CALC → 4: LinReg $(ax+b)$

LinReg $(ax+b)$: x List : L1

y List : L2



predicted value : if $x = k$

$\hat{y} = a(k) + b$

different between predicted value : a·d

slope : $m = \frac{y_2 - y_1}{x_2 - x_1}$

residual : observed value − predicted value

at point (x, y), $\hat{y} = ax + b$, then

residual $= y - \hat{y}$

coefficient of determination : $r^2 = \frac{\text{explained variation}}{\text{unexplained + explained variation}}$

if $r^2 = 0.xy$, then xy% of variation is explained by the least-square regression line

---

Ch 12 observed frequency : O

Expected frequency : $E_1 = np_1, E_2 = np_2, \ldots, E_n = np_n$

The chi-square statistic : $\chi^2 = \Sigma \frac{(O_i - E_i)^2}{E_i}$

df = K − 1 where K = number of categories

χ^2 GOF Test

$H_0 : p_1 = p_2 = \ldots = p_n = \frac{1}{K}$

H_1 : some or all the p_i differ from $\frac{1}{K}$

STAT → TEST → D: χ^2 GOF-Test (Observed : L1, Expected : L2, df)

Expected frequency : $E = \frac{\text{Row total} \cdot \text{Column total}}{\text{Grand total}}$

If $p < \alpha$, reject H_0. Enough evidence

if $p > \alpha$, do not reject H_0. Not enough evidence

χ^2 Test

H_0 : independent / same distribution

H_1 : not independent / not same

STAT → TEST → C: χ^2 Test (observed : [A], Expected : [B])

Ch 14 Grand mean : $\bar{\bar{x}}$ = average of all items of all samples

total number of samples : I

total number of items in all samples : N

Hypotheses for one-way ANOVA

$H_0 : \mu_1 = \mu_2 = \ldots = \mu_I$

H_1 : two or more μ_I are different

$SSTr = n_1(\bar{x}_1 - \bar{\bar{x}})^2 + n_2(\bar{x}_2 - \bar{\bar{x}})^2 + \ldots + n_I(\bar{x}_I - \bar{\bar{x}})^2$

$SSE = (n_1 - 1)S_1^2 + (n_2 - 1)S_2^2 + \ldots + (n_I - 1)S_I^2$

df for SSTr $= I - 1$

df for SSE $= N - I$

$MSTr = \frac{SSTr}{I-1}$, $MSE = \frac{SSE}{N-I}$

test statistic : $F = \frac{MSTr}{MSE}$

if $F > 1$, we reject H_0.

STAT → TESTS → H: ANOVA $(L_1, L_2, \ldots, L_I)$

if $p < \alpha$, we reject H_0.