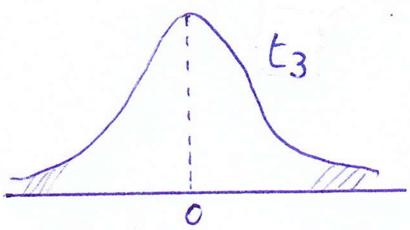
1) Velden bemesten of niet. Helft wel en helft niet bemesten.
Welke helft bemesten laten afhangen van kop of munt.

Naam: ....
Voornaam: ....

KLAD

2)



Zwaardere staarten → kurtosis > 3
Skewness = 0

3) Ongecorreleerd: $corr(u_i, X_i) = 0 \Leftrightarrow cov(u_i, X_i) = 0$ (teller)

$$cov(u_i, X_i) = E(u_i X_i) - E[u_i] \cdot E[X_i]$$
$$= E[E[u_i X_i | X_i]] - E[\underbrace{E[u_i | X_i]}_{0}] \cdot E[X_i]$$
$$= E[X_i \underbrace{E[u_i | X_i]}_{0}] = 0$$

4) $\mu = 0{,}1\%$
$\sigma = 2\%$

$$\underline{P(R \geq 0{,}1)} = P\left(\frac{R - 0{,}1}{2} \geq \frac{10 - 0{,}1}{2}\right)$$
$$= P(Z \geq 4{,}95)$$
$$= 1 - P(Z \leq 4{,}95) = 1 - 1 = \pm 0$$

5) Als n zeer groot, variantie steekproefgem. nt neer naar 0.
6) Als z'n variantie kleiner is
7) $E(Y) = P(Y=1)$ → verwachte waarden reduceren zich tot kansen wnr dependent var. dummy is.

8) Ja, blijft geldig, eigenschappen normale verdeling gebruiken maar t-verdeling mag omdat # vrijheidsgraden heel groot gaat worden & de normale verdeling gaat benaderen.
9) Consistentie: $r_{XY} \xrightarrow{P} \rho_{XY}$ $(n \to \infty)$
Voor n groot: $r_{XY} \approx \rho_{XY}$

10.1 Storingsterm & alle omitted var. zitten hierin

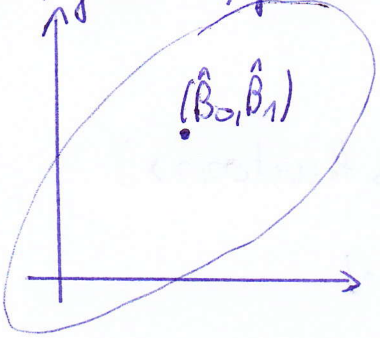
10.2 Vinden van m die ervoor zorgt dat: $\min \sum_{i=1}^{n} (Y_i - m)^2$

Hier: $\min \sum_{i=1}^{n} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2$

Deze schatters moeten de som van de residuen in kwadraat minimaliseren

10.3 Als n groot is, gaat de schatter ~ een normale verdeling

10.4



11) $Y_i = \beta_0 + \beta_1 X_i + u_i$

$E[Y_i | X_i] = \beta_0 + \beta_1 X_i$

$Var[Y_i | X_i] = Var[\underbrace{\beta_0 + \beta_1 X_i}_{\text{zoals een } c^{te}} + u_i | X_i] = Var[u_i | X_i]$

Assumptie homosked. => $Var[u_i | X_i] = \sigma^2 = c^{te}$

$E[(X - \mu) u_i] = E[E[(X - \mu_X) u_i | X_n]] = E[(X - \mu_X) \underbrace{E[u_i | X_n]}_{0}] = 0$

law of iterated expectations — want c^{te}

Als verwachte waarde = 0 => Var. = 2de moment

$E[(X - \mu_X)^2 u_i^2] = E[E[(X - \mu_X)^2 u_i^2 | X_i]] = E[(X - \mu_X)^2 \overbrace{E[u_i^2 | X_i]}^{\sigma^2 = c^{te}}]$

$= \sigma^2 \, Var(X_i) = Var[(X_i - \mu_X) u_i]$

Conclusie: $\sigma^2_{\hat{\beta}_1} = \frac{1}{n} \frac{\sigma^2}{Var(X_i)}$

12) (i) In klasse v. schatters die unbiased zijn.

(ii) Kleinste mogelijke variantie

13) $Y_i = \beta_0 + \beta_1 X_i + u_i$

(i) % migranten, gem. inkomen want verklaren beide Y en zijn gecorreleerd met klasgrootte

(ii) Parkeerruimte leraars

14) R^2 heeft als nadeel dat hij altijd groter is als je variabelen toevoegt.

15.1 X: $|t| = \left|\frac{30}{7{,}6}\right| = 3{,}95 > 2 \rightarrow$ significant

Age: $|t| = \left|\frac{12}{7{,}6}\right| = 1{,}58 < 2 \rightarrow$ niet significant

15.2 $]12 \pm 2 \cdot 7{,}6[=]-3{,}2\,;\, 27{,}2[$ → nul ligt in het BI dus we verwerpen de nulhypothese niet.

15.3 Als het # jaren werkervaring met 1 jaar stijgt, zal het maandelijks inkomen met € 30 stijgen, ceteris paribus en gemiddeld gezien

15.4 Als het # jaren werkervaring met 1 jaar stijgt, zal het maandelijks inkomen met ______

15.5 (i)
(ii) F-toets

16) $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$

$E[u_i | X_{1i}, X_{2i}] = 7X_{2i}$

Opl: $E[Y_i | X_{1,i}+1, X_{2i}] - E[Y_i | X_{1,i}, X_{2i}] = \beta_0 + (X_{1i}+1)\beta_1 + \beta_2 X_{2i} + \underbrace{E[u_i | X_{1i}+1, X_{2i}]}_{= 7X_{2i}}$

$- \beta_1 - X_{1i}\beta_1 - \beta_2 X_{2i} - \underbrace{E[u_i | X_{1i}, X_{2i}]}_{= 7X_{2i}}$

$= \beta_1 + (0) = \beta_1$

17) $\frac{1}{1 - \rho^2_{X_1, X_2}}$ Als perfecte multicollineariteit → $\frac{1}{1-1} = \frac{1}{0}$

18.1 $|t| = \left|\frac{-0{,}6}{0{,}04}\right| = 15 > 2 \rightarrow$ significant dus verkiezen kwadratische model

18.2 Ceteris paribus & gemiddeld gezien zal het verwachte maandelijkse inkomen stijgen met €12 als er een stijging is van de werkervaring met 1 jaar.

18.3 (i) $\hat{Y} = 905 + 12X - 0{,}6X^2 + \beta_3 \cdot \text{geslacht} + \beta_4(\text{geslacht} \cdot \#\text{jaren werkervaring})$

met geslacht $\begin{cases} 1 \text{ als vrouw} \\ 0 \text{ als man} \end{cases}$ bijvoorbeeld

19.1 $T=2$ $n=100$

$$Y_{i2} - Y_{i1} = \cancel{\beta_0} + \beta_1 X_{2i} + \lambda_2 + \cancel{\alpha_i} + u_{i,2}$$
$$- \cancel{\beta_0} - \beta_1 X_{1i} - \lambda_1 - \cancel{\alpha_i} - u_{i,1}$$
$$= \beta_1 (X_{2i} - X_{1i}) + \underbrace{(\lambda_2 - \lambda_1)}_{\text{constante term = intercept}} + \text{storingsterm}$$

19.2 $n-1 = 100-1 = 99$ dummy variabelen toevoegen

19.3 "die niet specifiek zijn voor een bepaalde tijd en die niet specifiek zijn voor een bepaalde entiteit"

19.4 - "Clustered standard errors" zijn robuust voor heteroskedasticiteit én voor correlatie in de tijd binnen 1 entiteit. Ze zijn dus "heteroskedasticity & autocorrelation-consistent" = HAC. Hiermee laten we autocorrelatie toe binnen een entiteit.
- Elke cluster bestaat uit een entiteit.

20.1 $|t| = \left|\frac{-0{,}04}{0{,}01}\right| = 4 > 2 \rightarrow$ significant (X_1)

$|t| = \left|\frac{0{,}07}{0{,}022}\right| = 3{,}18 > 2 \rightarrow$ significant (X_2)

X_1: een jongen heeft een grotere kans dat hij niet slaagt voor al zijn vakken dan een meisje

X_2: Hoe hoger het behaalde resultaat in het middelbaar, hoe meer kans dat de persoon slaagt op al zijn vakken.

20.2 $\phi(-3{,}2 - 0{,}04 \cdot 0 + 0{,}07 \cdot 70) - \phi(-3{,}2 - 0{,}04 \cdot 1 + 0{,}07 \cdot 70)$

$= \phi(1{,}7) - \phi(1{,}66) = 0{,}9554 - 0{,}9515 = 0{,}0039$

$\rightarrow$ Een meisje heeft 0,39 percentagepunten meer kans om te slagen op alles dan een jongen.

20.3 Meisje: $\phi(-3{,}2 - 0{,}04 \cdot 0 + 0{,}07 \cdot 71) - \phi(-3{,}2 - 0{,}04 \cdot 0 + 0{,}07 \cdot 70)$
$= \phi(1{,}77) - \phi(1{,}7) = 0{,}9616 - 0{,}9554 = 0{,}0062$

$\rightarrow$ Als bij een meisje het behaalde resultaat op het middelbaar stijgt met 1 punt, zal ze 0,62 procentpunten meer kans hebben om te slagen voor alle vakken.

Jongen: $\varphi(-3{,}2-0{,}04\cdot1+0{,}07\cdot71)-\varphi(-3{,}2-0{,}04\cdot1+0{,}07\cdot70)$

$= \varphi(1{,}73)-\varphi(1{,}66) = 0{,}9582-0{,}9525$

$= 0{,}0057$

...

20.4 Ben ik niet zeker van: $\underset{\beta_1}{0{,}04}$ want X_1 gaat van 0 naar 1

20.5 Bij een lineair model kunnen de voorspelde waarden groter zijn dan 1 of kleiner dan 0. Dit kan echter niet het geval zijn bij binaire Y-variabelen. Probit lost dit op door de waarden tussen 0 en 1 te houden door middel van de "cumulative probability distribution functions".

20.6 Bij het probit model gebruiken we de cumulatieve standaardnormale distributiefunctie (φ) en bij het logit model de cumulatieve standaard logistische distributiefunctie (F).